Advanced Synchronization Techniques for Continuous Phase Modulation

A Thesis
Presented to
The Academic Faculty

by

Qing Zhao

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Electrical and Computer Engineering
Georgia Institute of Technology
May 2006

Advanced Synchronization Techniques for Continuous Phase Modulation

Approved by:

Professor Gordon L. Stüber
Adviser and Committee Chair

Professor Eric I. Verriest
Department of ECE

Professor Geoffrey Li
Department of ECE

Professor Yajun Mei
Department of Mathematics

Professor John R. Barry
Department of ECE

Date Approved: 28 March 2006

To my parents, Chongding Zhao and Fulan Yin,

and my wife, Dawei

.

ACKNOWLEDGEMENTS

This thesis is dedicated to my parents, Chongdoing Zhao and Fulan Yin, and my dear wife, Dawei, for their patience and love through all these years. My parents did their best since my childhood to provide me the best educational opportunities. They have been a source of incessant support and encouragement in my life, which has been vital to all of my accomplishments. Dawei has walked with me through many difficult times during my graduate study. Without her love and understanding, it would have been much harder for me to get here today.

I shall always be grateful to my advisor, Professor Gordon L. Stüber, for his continuous guidance and support throughout my study at Georgia Tech. He does not give me a fish but he does all he can do to teach me how to fish. From my deep heart I appreciate his advising style, which gave me much freedom in research, help me build my knowledge scheme and motivated me to work independently. Without him, it is impossible for me to be such confident in my research area.

I am also thankful to Dr. Kenneth A. Stewart for his mentoring during my summer internships at Motorola Inc. in 2004 and 2005. He set up a role model for me showing the path how to succeed in the industry world. His success and influence might be one of primary reasons that drove me to join industry after graduation.

Moreover, I want to thank the committee members, Professor Geoffrey Li, Professor Eric I. Verriest, Professor John Barry and Professor Yajun Mei, for their valuable advices, suggestions and comments, which make this thesis substantially better.

Last but not least, I extend my thanks to my current and previous labmates in Wireless System Lab at Georgia Tech, including Dr. Jun Tan, Dr. Kihong Kim, Dr. Husang Kim, Dr. Apuva Mody, JoonBeon Kim, Galib Asadullah, Ping-Huang Chuang, Alenka Zajic, Chirag Patel and Heewon Kang, as well as my friends at the Sector of Mobile Device at Motorola Inc., including Dr. Yakun Sun, Dr. Guocong Song, Raja Bachu and Eoin Buckley.

TABLE OF CONTENTS

DEDICATION . . . . . **iii**

ACKNOWLEDGEMENTS . . . . . **iv**

LIST OF TABLES . . . . . **viii**

LIST OF FIGURES . . . . . **ix**

SUMMARY . . . . . **xii**

I INTRODUCTION **1**

1.1 Synchronization for Uncoded CPM 2

1.2 Iterative Synchronization for Coded CPM 6

1.3 Noncoherent Detection 10

1.4 Outline of the Thesis 11

II JOINT TIME AND PHASE RECOVERY FOR CPM **13**

2.1 Representation of CPM 13

2.1.1 Tilted-Phase Representation 14

2.1.2 Laurent's Expansion 15

2.1.3 Exponential Expansion 16

2.2 Joint Timing and Phase Recovery 18

2.2.1 Robust Time and Phase Recovery on Time-invariant Channels . . . 19

2.2.2 Robust Time and Phase Recovery on Time-variant Channels 22

2.3 Convergence Analysis of The Proposed Synchronizer 23

2.3.1 Asymptotic Behavior of The Proposed Synchronizer 23

2.3.2 Effect of Modeling Errors on Convergence 26

2.4 Numerical Results And Discussions 30

2.4.1 Synchronizer Design Issues 31

2.4.2 MSE Performance of the Proposed Synchronizer 34

2.4.3 Effect of Modeling Errors 37

2.4.4 BER Performance 39

III ITERATIVE PHASE SYNCHRONIZATION FOR SERIALLY CONCATENATED CPM 42

3.1 Innovations-Based MAP Estimation 42

3.1.1 A Near-Optimal Evaluation for Innovations-Based MAP Estimation 42

3.1.2 Suboptimal Algorithms for Innovations-Based MAP Estimation . . 45

3.2 CPM and RIC Codes 47

3.2.1 Design of RIC Codes 47

3.2.2 Rotational Invariance of CPM Trellis 48

3.3 Adaptive Iterative Phase Synchronization for SCCPM Systems 50

3.3.1 System Description 50

3.3.2 Adaptive Iterative Phase Synchronization 52

3.3.3 Inherent Limitations of Bidirectional A-SISO Algorithms 52

3.4 Numerical Results And Discussions 55

3.4.1 Performance of the FI-A-SISO Algorithms 55

3.4.2 Performance of the Forward-Only FI-A-SISO Algorithms 59

IV CONVERGENCE BEHAVIOR OF ITERATIVE PHASE SYNCHRONIZATION AND DETECTION 63

4.1 Summary of Known Results 63

4.1.1 Serially Concatenated Trellis-coded Modulation with Rotational Invariance 63

4.1.2 Adaptive Iterative Decoding Algorithms 64

4.1.3 Basics of Convergence Analysis 66

4.2 Convergence of Iterative Phase Synchronization 68

4.2.1 A Simplified Phase Model 68

4.2.2 Synchronization without Cycle Slipping 69

4.2.3 Synchronization with Cycle Slipping 72

4.3 The Effect of Doppler Spread on Mean Time to Cycle Slipping 79

4.4 Numerical Results And Discussions 82

4.4.1 Effect Of Decision Depth for SEP-based A-SISO Algorithms 83

4.4.2 Selection of Iteration-stopping Criterion 84

4.4.3 Effects of Interleaver Size 87

4.4.4 Performance of Different SCCPM Systems 89

V TURBO SYNCHRONIZATION FOR SERIALLY CONCATENATED CPM . . . 91

5.1 Turbo Synchronization for SCCPM . . . 91

5.1.1 System Description . . . 91

5.1.2 The A-SISO CPM Demodulator . . . 93

5.2 Numerical Results And Discussions . . . 94

5.2.1 BER Performance of Turbo Synchronization . . . 95

5.2.2 Effects of Incorrect Modeling . . . 97

VI NONCOHERENT SEQUENCE DETECTION ON RAYLEIGH FADING CHANNELS . . . 99

6.1 System Model and Problem Formulation . . . 99

6.1.1 Channel Model . . . 100

6.1.2 Coherent and Noncoherent Sequence Detection . . . 101

6.1.3 Error Probability of Coherent Sequence Detection . . . 104

6.2 Noncoherent Sequence Detection . . . 106

6.2.1 Distribution of Hermitian Quadratic Form . . . 106

6.2.2 On Time-invariant Channels . . . 107

6.2.3 On Time-variant Channels . . . 110

6.3 A Hybrid Scheme on Time-variant Channels . . . 114

VII CONCLUSIONS AND FUTURE WORK . . . 118

7.1 Conclusions . . . 118

7.2 Future Work . . . 120

APPENDIX A — SMOOTHING ESTIMATION . . . 121

APPENDIX B — EQUIVALENCE OF (59) AND (60) . . . 123

APPENDIX C — PROOF OF THE EQUATION (62) . . . 125

APPENDIX D — PROOF OF THE EQUATION (127) . . . 127

APPENDIX E — PROOF OF THE EQUATION (128) . . . 128

APPENDIX F — PROOF OF THE EQUATIONS (139) AND (140) . 129

REFERENCES . . . 131

VITA . . . 140

LIST OF TABLES

Table 1 Performance of different CPM signal exponential expansions. 17

Table 2 Performance of the 3-dimentional nOEEs and constrained nOEEs. . . . 32

Table 3 Performances of different iteration-stopping criteria.[a] 86

Table 4 The practical values of the coefficient ρ on time-variant fading channels. . 100

LIST OF FIGURES

Figure 1 Normalized MSE performance of the proposed synchronizer with double- and triple-initialized Kalman filtering. The constrained nSnOEE with 3 pulses is used. The normalized time variance is defined as $\mathrm{var}(\hat{\tau}_k)/T_s^2$. . . 33

Figure 2 Normalized MSE performance of the proposed synchronizer with constrained SnOEE and nSnOEE for a 4-ary-3RC CPM system with $h_{\mathrm{idx}} = 1/2$. The training length is $N_u = 200$. A triple-initialized Kalman filter with $N_1 = 40$ and $N_2 = 80$ is employed. The normalized time and phase variance are defined as $\mathrm{var}(\hat{\tau}_k)/T_s^2$ and $\mathrm{var}(\hat{\phi}_k)/\pi^2$. 34

Figure 3 Normalized MSE performance of the proposed synchronizer with constrained SnOEE and nSnOEE for a 4-ary-3RC CPM system with $h_{\mathrm{idx}} = 1/2$. The training length is $N_u = 20$. A triple-initialized Kalman filter with $N_1 = 5$ and $N_2 = 12$ is employed. The normalized time and phase variance are defined as $\mathrm{var}(\hat{\tau}_k)/T_s^2$ and $\mathrm{var}(\hat{\phi}_k)/\pi^2$. 35

Figure 4 Normalized MSE performance of the proposed synchronizer with constrained SnOEE and nSnOEE for GMSK with $BT_s = 0.30$. The training length is $N_u = 200$. A triple-initialized Kalman filter with $N_1 = 40$ and $N_2 = 80$ is employed. The normalized time and phase variance are defined as $\mathrm{var}(\hat{\tau}_k)/T_s^2$ and $\mathrm{var}(\hat{\phi}_k)/\pi^2$. 36

Figure 5 Normalized timing MSE performance of the proposed synchronizer with incorrect modeling errors for a 4-ary-3RC CPM system with $h_{\mathrm{idx}} = 1/2$. A triple-initialized Kalman filter with $N_1 = 40$ and $N_2 = 80$ is employed. The normalized time variance is defined as $\mathrm{var}(\hat{\tau}_k)/T_s^2$. 37

Figure 6 Convergence performance of the proposed synchronizer with incorrect modeling errors for a 4-ary-3RC CPM system with $h_{\mathrm{idx}} = 1/2$ on a time-invariant channel with $E_s/N_o = 20\,\mathrm{dB}$. A triple-initialized Kalman filter with $N_1 = 40$ and $N_2 = 80$ is employed. The normalized time variance is defined as $\mathrm{var}(\hat{\tau}_k)/T_s^2$. 38

Figure 7 Estimate trajectories of the proposed synchronizer with correct and incorrect modeling errors with $E_s/N_o = 15\,\mathrm{dB}$. Triple-initialization with $N_1 = 40$ and $N_2 = 80$. 39

Figure 8 BER performance of the proposed receiver for a 4-ary-3RC CPM system with $h_{\mathrm{idx}} = 1/2$. The nSnOEE-based FEP is used. A triple-initialized Kalman filter with $N_1 = 40$ and $N_2 = 80$ is employed. 40

Figure 9 Trellis diagram of RIC codes. 48

Figure 10 Rotational invariance of CPM trellis. 49

Figure 11 Transmitter and receiver structure for SCCPM systems. 51

Figure 12 Survivor paths within MSK trellis. 53

Figure 13 MSE of phase estimates for the FI-A-SISO algorithms. An SCMSK system operating on an AWGN channel with a phase step is considered. The iterative receiver stops after 2 iterations. The interleaver size is $N_u = 500$. 56

Figure 14 BER performance of the FI-A-SISO algorithms for both noniterative and iterative receivers. For iterative receivers the final decisions are made after 10 iterations. The interleaver size is $N_u = 500$. 57

Figure 15 BER performance of the FI-A-SISO algorithms for an AWGN channel with a phase ramp. The corresponding frequency offset is $0.05/T_s$. A SCMSK system is considered. The decisions are made after 10 iterations. The interleaver size is $N_u = 500$. 58

Figure 16 BER performance of the *forward-only* FI-A-SISO algorithms for an AWGN channel with different random phase jitters. A SCMSK system is considered. The decisions are made after 10 iterations. The interleaver size is $N_u = 2048$. 60

Figure 17 BER performance of the *forward-only* FI-A-SISO algorithms for an AWGN channel with a phase ramp. The corresponding frequency offset is $0.05/T_s$. A SCMSK system is considered. The decisions are made after 10 iterations. The interleaver size is $N_u = 2048$. 61

Figure 18 BER performance of the *forward-only* FI-A-SISO algorithms for a correlated Rayleigh fading channel. The normalized maximum Doppler shift is $f_m T_s = 0.001$. A SCMSK system is considered. The decisions are made after 5 iterations. The interleaver size is $N_u = 8192$. 62

Figure 19 Transmitter structure for SCTCM. 64

Figure 20 Two examples of RI-TCM systems 65

Figure 21 Iterative demodulation/decoding scheme for SCTCM. 66

Figure 22 EXIT chart for SCMSK. $N_x = 100000$. 67

Figure 23 Performance loss owing to phase disturbance. $E_b/N_o = 3.0\,\mathrm{dB}$ 71

Figure 24 EXIT chart for SCDBPSK with phase disturbance. $N_x = 100000$. 72

Figure 25 DBPSK trellis alternation owing to the occurrence of cycle slipping. . . . 73

Figure 26 Evolution of mutual information at inner decoder of SCDBPSK with Type-I cycle slipping. $E_b/N_o = 3\mathrm{dB}$. $N_x = 100000$. 74

Figure 27 Evolution of mutual information at inner decoder of SCDBPSK with Type-II cycle slipping. $N_x = 100000$ 76

Figure 28 Structure of phase-locked loops. 80

Figure 29 Mean time to cycle slipping over AWGN and time-selective fading channels. 82

Figure 30 Effect of decision depth on the SEP-based A-SISO algorithms. SCMSK is considered with $N_x = 2048$. 83

Figure 31 Effect of interleaver size over time-selective fading channels. The SEP-based A-SISO demodulator with the *SNR* stopping criterion is employed. $E_b/N_o = 10\,\text{dB}$. 87

Figure 32 Performances of different SCCPM systems on a fading channel with $f_m T_s = 0.001$. The SEP-based A-SISO demodulator with the *SNR* stopping criterion is employed. $N_x = 4096$. 88

Figure 33 Performances of different SCCPM systems on a fading channel with $f_m T_s = 0.01$. The SEP-based A-SISO demodulator with the *SNR* stopping criterion is employed. $N_x = 2048$. 90

Figure 34 BER performance of the proposed turbo synchronizer. 96

Figure 35 The effects of incorrect channel modeling. 98

Figure 36 Performance of MSK noncoherent and coherent detection on time-invariant fading channels. 109

Figure 37 Optimal modulation indices of full-response CPM with rectangular frequency shaping pulse for noncoherent and coherent detection on time-invariant fading channels. 110

Figure 38 Performance of MSK noncoherent and coherent detection on time-variant fading channels. 112

Figure 39 Performance of MSK noncoherent and coherent detection on time-variant fading channels with $E_s/N_o = 20\,\text{dB}$. 113

Figure 40 Performance of MSK noncoherent and coherent detection on time-variant fading channels with $E_s/N_o = 30\,\text{dB}$. 114

Figure 41 Power efficiency of noncoherent and coherent detection given service quality. 115

Figure 42 Optimal observation width of noncoherent detectors given service quality. 116

SUMMARY

The objective of this research work is to develop reliable and power-efficient synchronization algorithms for continuous phase modulation (CPM). CPM is a bandwidth and power efficient signaling scheme suitable for wireless and mobile communications. Due to phase continuity and constant modulus, CPM signals exhibit a narrow spectral mainlobe and low sidelobes, and allow nonlinear power amplifiers to operate efficiently. Binary CPM schemes have been widely used in many commercial and military systems. CPM with multi-level symbol inputs, i.e., $\mathcal{M}$-ary CPM, can achieve a higher data rate than binary CPM. However, the use of $\mathcal{M}$-ary CPM has been limited due to receiver complexity and synchronization problems. In the last decade, serially concatenated CPM (SCCPM) has drawn more attention since this "turbo-like" coded scheme can achieve near Shannon-limit performance by performing iterative demodulation/decoding. Note that SCCPM typically operates at a low signal-to-noise ratio, which makes reliable and power-efficient synchronization more challenging.

In this thesis, we propose a novel timing and phase recovery technique for CPM, based on Kalman filtering and an approximate representation of CPM signals with nonorthogonal exponential expansions. Compared to existing maximum-likelihood estimators, the proposed data-aided synchronizer can achieve a better acquisition performance when a preamble is short or channel model errors are present. Extending this synchronizer to its decision-directed version is also possible. However, to ensure reliable parameter tracking, we entail to pilot-symbols and periodically perform pilot-assisted estimation during the decision-directed synchronization. The resulting mixed synchronizer exhibits an excellent performance in tracking time-variant phase and time errors.

We also propose a novel adaptive soft-input soft-output (A-SISO) module for iterative detection with parameter uncertainty. In contrast to the existing A-SISO algorithms using linear prediction, the parameter estimation in the proposed structure is performed in a

more general least-squares sense. Based on this scheme, a family of fixed-interval A-SISO algorithms are utilized to implement blind iterative phase synchronization for SCCPM. By exploiting inherent rotational invariance of CPM trellis structure, it is shown that the proposed iterative receivers are robust over channels of high and random phase dynamics. Moreover, the convergence characteristics of iterative phase synchronization and detection are analyzed by means of density evolution. Particularly, an oscillatory convergence behavior is observed when cycle slips occur during phase tracking. In order to reduce performance degradation due to this convergence fluctuation, design issues, including delay depth of the proposed algorithms, iteration-stopping criteria and interleaver size, are also discussed.

Finally, for completeness of the study on phase synchronization, we investigate the error probability performance of noncoherently detected full-response CPM, which does not require channel (or phase) estimation. This analysis explicitly indicates how the performance of noncoherent detection is affected by detector complexity, fading rate and operating signal-to-noise ratio. Furthermore, using this result, we propose a hybrid system which switches between a noncoherent detector and a coherent detector with imperfect channel estimates, in order to optimize the overall system performance in terms of service quality, receiver complexity and power efficiency.

CHAPTER I

INTRODUCTION

Continuous phase modulation (CPM) is a bandwidth and power efficient signaling scheme suitable for wireless and mobile communications [1]-[4]. Compared to linear modulation schemes with abrupt phase changes, such as $\mathcal{M}$-ary phase-shift-keying (PSK) and quadrature amplitude modulation (QAM), CPM signals exhibit a narrower spectral mainlobe and lower sidelobes due to their phase continuity. The resulting compact spectrum enables efficient CPM signal multiplexing. CPM waveforms also have a constant envelope, which allows nonlinear power amplifiers to operate efficiently at saturation. Additional CPM power efficiency enhancement also results from the inherent memory structure of CPM signals, i.e., phase continuity, which increases minimum Euclidean distance between possible transmitting signals. Binary CPM schemes have been widely used in commercial systems, such as Gaussian frequency-shift-keying (GFSK) in Bluetooth and Gaussian minimum-shift-keying (GMSK) in Global System for Mobile communication (GSM) [5]. CPM with multi-level symbol inputs, i.e., $\mathcal{M}$-ary CPM, can achieve a higher data rate than binary CPM. However, the use of $\mathcal{M}$-ary CPM has been limited due to receiver complexity and synchronization problems, including timing recovery, frequency offset and carrier phase estimation [6]. In the last decade, serially concatenated CPM (SCCPM) has drawn more attention since this "turbo-like" coded scheme can achieve near Shannon-limit performance by performing iterative demodulation/decoding [7],[8]. Note that SCCPM typically operates at a low signal-to-noise ratio (SNR), which makes reliable and power-efficient synchronization more challenging. In this thesis, we design and analyze synchronization algorithms for both uncoded and coded CPM systems. For completeness, we also investigate noncoherent CPM demodulation, which does not require carrier phase recovery.

1.1 Synchronization for Uncoded CPM

Synchronization can be generalized as a typical parameter estimation problem, which can be tackled by using data-aided (DA), decision-directed (DD) or non-data-aided (NDA) techniques [6]. Using the DA method, a parameter estimator is driven by known data symbols, e.g., a training sequence placed at the beginning of data transmission or a set of pilot symbols periodically inserted into a data information sequence. In the DD method, no data information is available at receiver and a set of tentative data decisions are used to aid parameter estimation in the same way as the training or pilot symbols do. For a NDA synchronizer, neither known data symbols or tentative decisions are needed and unknown parameters are estimated by exploiting the high-order statistics or other special nonlinear features of received signals. Normally, among them, the DA method can achieve the best estimation performance, including hangup-free, fast convergence and Cramer-Rao-Bound-approaching mean-square-error (MSE), although they yield the lowest data rate or power efficiency due to training overhead. Apparently, a DD or NDA estimator allows for a higher data rate than a DA synchronizer; however, a DD estimator is usually more complex and might experience hangup problems in certain circumstances, and the MSE performance of a NDA estimator is poor. In practice, the combination of these methods might be considered to improve overall system performance.

In most practical applications, a timing error and a frequency offset vary much more slowly than a carrier phase. In particular, the frequency offset can thus be presented by a phase ramp over one or a few data frames, which can always be recovered by a well-designed phase estimator, e.g., a second-order phase lock loop (PLL) [9]. In this thesis, for simplicity of design and analysis, we ignore frequency offset estimation, hereafter, and consider only time and phase recovery, which can be either processed individually in a sequential order or jointly at the same time. Typically, separate parameter estimation is less complex; however, it requires that the estimation of the first parameter must be robust when the other is unknown.

In the scenario of separate estimation for CPM, the carrier phase with a large initial error is first acquired with clockless NDA phase estimation [6]. Then CPM timing recovery

is performed with the aid of the previously obtained coarse phase estimate. Once the time error is acquired, fine clock-aided phase acquisition is further processed to provide a more accurate phase estimate for followed coherent CPM demodulation. Then another round of fine timing recovery can also be executed if necessary. On the other hand, it is also possible to first perform clock recovery without the knowledge of the carrier phase [10], and then the rest of the synchronization procedure is similar to the case in which the phase offset is first estimated.

Both clock-aided phase acquisition or phase-aided timing recovery can be implemented in either a DD, DA or NDA fashion, depending on the choice of an application. The fundamental design rule for single parameter estimation is similar and can be summarized as follows. The DA method derived from a maximum-likelihood (ML) criterion is an *adhoc* approach requiring a feedback or closed-loop structure, and extending it to its DD counterpart is straightforward except that the DD estimation must be processed jointly with CPM demodulation, leading to a complex receiver. By taking average on data information, the DA-ML estimation simply reduces to the NDA-ML estimation, which can be implemented in either a feedback or a feedforward fashion although both suffer from a poor MSE performance. However, the NDA-ML feedforward estimator spends less time to achieve reliable acquisition than other feedback schemes. Practical ML-based estimation algorithms can be found in [6], [11] and [12] for CPM phase recovery and [6], [13], [14] and [15] for CPM timing recovery. Moreover, the non-ML-based estimation exploiting special features of CPM signals is also available and is usually designed in a NDA feedforward fashion. A robust non-ML phase estimator can be found in [6], while two non-ML time error estimators were proposed in [16] and [17] though both work well only for CPM modulation of a minimum-shift-keying (MSK) signal type[1].

The primary difference between the ML-based individual phase acquisition and time recovery is that an additional set of *derivative* matched filters is required for any-type timing recovery method when generating appropriate sufficient statistics, because the log-likelihood function of CPM signals is nonlinear in a time offset. To reduce the complexity,

[1]The MSK-*type* modulation refers to binary CPM with the modulation index $h = 1/2$.

two possible approaches are considered in the literature [6]: 1) approximating *derivative* matched filtering with finite difference of outputs from standard matched filters sampled at the symbol rate; 2) linearizing the nonlinear log-likelihood function directly by assuming a sufficiently small timing error. Both methods perform poorly, especially when the initial timing error is significant or when the $\mathcal{M}$-ary partial-response CPM is considered.

For joint CPM time and phase recovery, the ML-based DA method has been well-studied since the invention of CPM [4], [18], [19]. However, extending these DA synchronizers to their DD counterparts is not trivial as in single parameter estimation, because multiple stable local optima exist in the likelihood function. It has been reported in [20] that the ML-DD joint time and phase recovery cannot synchronize $\mathcal{M}$-ary partial-response CPM signals due to false locks in a timing loop, although it performs well for other CPM signals. In order to overcome this dilemma, a false lock detector and an auxiliary acquisition unit must be carefully devised to detect and push a ML-DD synchronizer away from false equilibria [20], [21]. These additional processes increase receiver complexity, and a long acquisition time is normally required to relock synchronization.

Another impediment for the widespread use of CPM is receiver complexity, which is determined not only by the complexity of a demodulator but also by the complexity of the front-end processor (FEP) that is used to generate sufficient statistics for parameter estimation and data detection. An optimal coherent CPM demodulator requires a maximum likelihood sequence detector, which is especially complex for $\mathcal{M}$-ary partial-response CPM [4]. Suboptimal detectors, such as linear minimum mean-square-error (MMSE) detectors [22] and reduced-state sequence detectors [23], [24], have been devised to reduce the complexity. To reduce the complexity of the FEP, one can approximate CPM signals with a small set of basis functions. The accuracy and efficiency of the signal approximation depends on the selection of basis functions. Laurent's expansion was the earliest attempt toward this end, although the corresponding basis functions usually have a complicated partial-response form [25]. Other sets of basis functions having simpler forms were subsequently proposed, including exponential functions [26], sinc$(\cdot)$ functions [27] and Walsh functions [28]. However, all of these approaches are near-optimal only in preserving the

minimum distance property of CPM ML detection, but not in approximating CPM signals themselves.

In this thesis, we propose a novel DA joint timing and phase recovery method that requires a simple FEP and is able to achieve fast and reliable parameter acquisition for all CPM systems. Using the fact that timing information can be easily extracted from exponential waveforms without requiring *derivative* matched filtering, we devise a new set of exponential basis functions that is more accurate and efficient in approximating CPM signals than the exponential bases used in [26]. Based on the proposed signal expansion, we build a discrete-time state-space model for received CPM signals. Extracting timing and phase information directly from this model incurs a nonlinear problem that requires extended Kalman filtering, which usually suffers from false locks and divergence. Instead of using this direct method, we first estimate a state vector that conveys the complete timing and phase information, and then update the desired parameter estimates from the estimate of this state vector. A similar approach has been well-studied for phase synchronization [29]. However, its extension to joint timing and phase synchronization is non-trivial, being feasible only when modulated signals can be accurately approximated with exponential expansions. Compared to the traditional ML-DA estimators with *derivative* matched filters [20], the proposed synchronizer has a less complex FEP and can achieve better acquisition performance, especially with a short length preamble. Based on the concept of uniform observability, we also show that the proposed synchronizer is *stable* on time-variant channels and *exponentially stable* on time-invariant channels.[2] Moreover, several modeling parameters, such as the variance of channel noise, the variances of parameter jitters and the initial state error covariance, are required to design the proposed synchronizer. However, they are known only approximately in practice. Hence, we also investigate the effect of the modeling errors on the synchronizer convergence. The selection of suboptimal exponential bases and the design of triple-initialized Kalman filters are also discussed. Our simulation results show that the proposed synchronizer has an excellent performance in acquiring and tracking both

[2] An estimator is *stable* if its estimate error variance is bounded; an estimator is *exponentially stable* if its estimate error variance converges exponentially to zero [30].

the time shift and phase offset on either time-invariant or time-variant channels.

The extension of the proposed DA synchronizer to its decision-directed version is also possible. However, to effectively prevent the occurrence of false locks, we insert pilot-symbols into a data sequence and periodically perform pilot-assisted estimation during the DD synchronization to ensure a timing loop is correctly locked. Based on the combination of both DA and DD concepts, this mixed method exhibits an excellent performance in tracking time-variant phase and time errors.

1.2 Iterative Synchronization for Coded CPM

The serial concatenation of a convolutional code, a random interleaver and a CPM modulator yields a type of serially concatenated convolutional codes, called a "turbo-like" code [31], [32]. As the inner code of this turbo-coded system, CPM has a recursive trellis structure and, thus, a near Shannon-limit performance can be achieved by performing iterative demodulation/decoding [7], [33]. However, this system typically operates at a low SNR and, therefore, reliable synchronization is challenging, especially in burst mode transmission allowing only a short preamble. Furthermore, for a turbo-coded scheme, using a longer data frame yields a larger interleaving gain. Nevertheless, the consequence using a long data frame is that one cannot simply assume a parameter is constant over an entire frame even when it has low dynamics. In such a case, parameter time-variations must be tracked approximately during iterative detection. In most practical applications, the carrier phase varies more rapidly than the timing error. Ignoring phase time-variations over data transmission may significantly degrade the performance of CPM coherent detection. Although not common, the timing dynamics also need to be tracked in some extreme circumstances. In this thesis, we first focus on iterative phase synchronization and then move to turbo synchronization for joint time and phase recovery.

The pilot-assisted CPM demodulation methods proposed in [34] and [35] are robust in acquiring and tracking a time-variant carrier phase, but their power efficiency is poor on fast fading channels because of a high training overhead. To improve power efficiency while

retaining good synchronization performance, joint phase estimation and CPM demodulation is a promising alternative. An early attempt toward this end was presented in [11], where linear prediction of flat fading was processed along with trellis-based CPM detection in a DD manner. The basic idea therein was later developed into a general solution, i.e., per-survivor processing (PSP), for hard-output sequence detection with unknown parameters [36], [37]. Unfortunately, PSP is not directly suitable for the soft-input soft-output (SISO) maximum *a posteriori* (MAP) symbol detectors [38], [39] , which are typically used in iterative receivers. For this reason, a family of adaptive SISO (A-SISO) algorithms were proposed in [40] for MAP-*type* iterative detection, based on an extension of PSP. Also, in [41], a set of *fixed-interval* A-SISO (FI-A-SISO) algorithms are developed for carrier phase tracking in turbo-coded quadrature phase-shift keying (QPSK) modulated systems. More recently, in [42], a set of *sliding-window* A-SISO (SW-A-SISO) algorithms were applied to SCCPM systems over unknown flat fading channels. Other applications of A-SISO algorithms for timing recovery and channel state estimation can also be found in [43]-[47]. In the literature, the A-SISO algorithms fall into two categories, the *forward-only* and *forward-backward* A-SISO algorithms. Within the *forward-only* A-SISO algorithms, parameter prediction is processed only in the forward direction; for the *forward-backward* A-SISO algorithms, both forward and backward linear prediction are required [48]. Simulation results in [41] and [42] show that the performance of the *forward-only* A-SISO algorithms is inferior to but closely approaches that of the *forward-backward* A-SISO algorithms.

In this thesis, we propose a novel A-SISO module that combines MAP symbol detection with innovations-based parameter estimation. Furthermore, a class of suboptimal reduced-complexity A-SISO algorithms are derived in a PSP fashion. In contrast to the existing A-SISO algorithms that use linear predicted estimates, the proposed scheme employs a set of more general linear least-squares estimates (LLSEs), including smoothed, filtered and predicted estimates, in metric recursions. The "innovations-based" notion herein follows from the fact that the LLSEs can be recursively updated by using a set of innovations processes. Note that the "innovations-based" estimation scheme in both [49] and [50] is essentially equivalent to linear prediction, which is one special case of the proposed structure.

Based on the proposed methodology, a family of FI-A-SISO algorithms are developed for iterative phase synchronization on SCCPM receivers. Compared to the SW-A-SISO algorithms in [42], the FI-A-SISO algorithms can achieve lower bit-error-rate (BER) with less complexity, but require more memory. The *forward-only* FI-A-SISO algorithms are shown, surprisingly, to outperform the *forward-backward* FI-A-SISO algorithms in acquiring and tracking a time-variant phase, especially at a low SNR. The reason has to do with rotational invariance of CPM signals. This result can be extended to any rotationally invariant convolutional (RIC) coded system.

RIC codes were invented by Wei in 1984 [51, 52]. By careful design, RIC codes can tolerate discrete phase ambiguities by ensuring that a phase-rotated code sequence is also a valid code sequence and carries the same information bits as the original code sequence. This property is particularly valuable for phase synchronization, since the performance degradation due to false lock, cycle slip and hangup can be easily avoided or reduced. In his pioneered work, Wei proposed the basic concepts and design rules for RIC codes by viewing them as as a special realization of trellis-coded modulation (TCM). Many new RIC codes were later found by using either synthesis methods [53, 54] or exhaustive searches [55]. The typical examples of RIC codes include differential codes and CPM. In [56], based on graph theory, a theoretical framework underlying RIC codes was presented together with necessary and sufficient conditions for a code to be rotationally invariant. It was also shown in [57] that a sufficient condition for a turbo-like code to be rotationally invariant is that its inner code is rotationally invariant. More general design rules and performance analysis for turbo-like codes with rotational invariance can be found in [58].

For SCCPM operating at a low SNR, reliable phase synchronization is a challenging task. When radio channels have high and random dynamics, accurate phase tracking is far more difficult. In these circumstances, phase synchronization is subject to frequent cycle slipping, false-lock and hang-up [9]. Although SCCPM has a strong robustness to these phase ambiguities because of rotational invariance of its inner code, the effect of imperfect phase synchronization on the convergence of iterative demodulation/decoding is still of

interest. It was first reported in [59] that iterative detection of rotationally invariant turbo-like codes does not consistently converge through iterations when discrete phase rotations exist. When carrier phase is rapidly time-varying, an oscillatory convergence behavior was observed at iterative receivers that use DD phase-tracking. This thesis provides a semi-theoretical analysis on this surprising convergence behavior. To facilitate our analysis, we first present a simplified phase model to differentiate the effects of small phase disturbances and discrete phase ambiguities. With this model we show that: 1) the phase disturbance does not cause convergence fluctuations and its effect can be simply viewed as an equivalent power reduction; 2) the oscillatory convergence behavior results from phase ambiguities in phase-tracking.

As indicated earlier, the practical DD algorithms in the context of MAP symbol detection (MAPSD) are usually implemented in a PSP fashion, where multiple tentative parameter estimates are generated at each symbol interval and used for metric recursions. The single-estimator processing (SEP) [60] retaining only one parameter estimate is thought of as a simplified PSP technique. It was shown in [41] that the PSP-based A-SISO algorithms outperform their SEP-based counterparts on both phase tracking and detection when cycle slipping, false-lock, and hang-up do not occur. However, our analysis implies that in the presence of cycle slipping, false-lock and hang-up, the SEP-based A-SISO algorithms can achieve a better convergence performance. This conjecture is confirmed by simulation results. Due to crucial effects of cycle slipping on convergence of iterative phase synchronization, we also provide an approximate method to evaluate the mean time of cycle slipping on fading channels. Moreover, design issues such as decision depth in the SEP method, iteration-stopping criteria, and effects of interleaver size on convergence, are discussed.

In this thesis, we also provide a turbo synchronization scheme to implement joint time and phase recovery for SCCPM. To prevent the false locks usually appearing in a timing loop, the preamble and pilot symbols are also used together with tentative decisions to facilitate reliable synchronization.

1.3 Noncoherent Detection

Maximum likelihood sequence detection can be implemented either coherently or noncoherently once the knowledge of timing and frequency offset is available. Coherent detection requires an estimate of the carrier phase, more strictly speaking, an estimate of a complex channel gain on flat fading channels[3], while noncoherent detection does not. The performance of a channel estimator generally degrades with increasing channel dynamics, and will eventually limit the performance of coherent detection. Noncoherent detection was originally proposed for time-invariant channels, where its BER performance closely approaches that of coherent detection. When the channel is time-varying, the performance of noncoherent detection degrades. A comparison between noncoherent and coherent detection on time-variant fading channels with optimal/suboptimal channel estimation, in terms of complexity and performance, remains an open problem.

Past research has evaluated coherent detection with imperfect channel knowledge using BER, cutoff rate or mutual information. In [61]-[65], the effects of channel estimation errors on BER were investigated for different scenarios, including frequency selective and non-selective Rayleigh and Rician fading channels, antenna diversity, and uncoded or coded modulation. Moreover, data-aided training strategies that optimized pilot placement and power allocation were proposed for transmission on time-variant flat fading channels, in the sense of minimizing uncoded BER in [66] or cutoff rate in [67]. Considering imperfect channel knowledge, an information-theoretic analysis on mutual information has also been provided for i.i.d multiple-input multiple-output flat fading channels in [68], i.i.d multipath block fading channels in [69], and temporally correlated flat fading channels in [70]. From different perspectives, the above literature consistently shows that the performance gap between coherent detection with and without ideal channel information is irreducible and dependent on the fading rate.

Research on noncoherent detection was initiated with robust receiver design for continuous phase frequency shift-keying (CPFSK) modulation in the 1960's [71], [72], and

[3]To align with the terminology referred in the literature, we use "channel" in stead of "phase" in the remainder of the thesis when comparing coherent and noncoherent detection.

then was extended to general TCM in the 1990's [74], [75], and most recently to capacity-approaching serial concatenated coded systems [33], [76]. The optimal input distribution, code, and constellation design for efficient noncoherent detection can be found in [77]-[79]. Previous simulation results have shown that the performance of noncoherent detection on time-invariant fading channels consistently improves with increasing detector complexity [80]-[83], while on time-variant channels the performance varies depending on both the detector complexity and the channel fading rate [84], [85]. Meanwhile, a significant effort has also been undertaken to derive error probabilities for noncoherent detection on fading channels; however, most existing results contain at least one integral function and defy a simple solution. Moreover, the existing analytical results fail to explain how the detector performance is affected by detector complexity, fading rate, and operating SNR.

In this thesis, we compare noncoherent and coherent detection on Rayleigh flat fading channels. By using a first-order Gauss-Markov channel model, we first derive a simple, closed-form, expression for error probabilities of noncoherently detected full-response CPM and differential phase-shift-keying (DPSK) modulation, on both time-invariant and time-variant fading channels. Our results explicitly show how to design optimal noncoherent detectors under different channel conditions. One of our more surprising conclusions is that a less complex noncoherent detector may achieve a better BER performance on fast fading channels than a more complex detector. Based on our analysis, we propose a hybrid system which switches modes between noncoherent detection and coherent detection with optimal/suboptimal channel estimation, in order to optimize the overall system performance, in terms of service quality, receiver complexity and power efficiency. The design issues, such as the mode switching criterion and the selection of a noncoherent detector with appropriate complexity, are also discussed.

1.4 Outline of the Thesis

The remainder of this thesis is organized as follows. In Chapter II, we propose a novel joint time and phase synchronizer for uncoded CPM, which requires a less complex front-end

processor and converges more rapidly as compared to existing ML-based estimators. In Chapter III, we devise a family of adaptive iterative algorithms to implement joint phase synchronization and CPM demodulation for coded CPM, i.e., SCCPM. By revisiting the rotational invariance of CPM signals, we further investigate the convergence behavior of the proposed iterative receivers with imperfect phase estimation in Chapter IV. Integrating the concepts of both data-aided and decision directed techniques, in Chapter V, we propose a turbo synchronization scheme for joint timing and phase recovery at iterative SCCPM receiver. In Chapter VI, we study the performance of noncoherent sequence detection for uncoded CPM. Chapter VII finally summarizes the contributions of the thesis and suggests the direction of future work in the related areas.

CHAPTER II

JOINT TIME AND PHASE RECOVERY FOR CPM

A brief review on existing representations of CPM signals is provided in this chapter, and then a new representation format, called nonorthogonal exponential expansion (nOEE), is proposed[1] . Using nOEE, we devise a novel data-aided joint time and phase recovery method based on Kalman filtering. Compared to existing ML-based estimators, the proposed synchronizer requires a less complex FEP and can achieve much more reliable acquisition performance with a short preamble. The asymptotic stability and convergence of the proposed synchronizer is also analyzed, including the effect of statistical channel modeling errors on the convergence characteristics. Moreover, the selection of suboptimal nOEEs and the design of triple-initialized Kalman filters are discussed. Both theoretical and simulation results show that the proposed synchronizer is robust in acquiring and tracking both time shift and phase offset on either time-invariant or time-variant channels. The main results in this chapter can also be found in [86].

2.1 Representation of CPM

The baseband complex envelope of CPM signals is

$$x(t,\mathbf{u}_1^k)=\sqrt{\frac{E_s}{T_s}}\ \exp\left\{j\varphi(t,\mathbf{u}_1^k)\right\},\quad (k-1)T_s\leqslant t\leqslant kT_s,\ \ k=1,\cdots,N_u, \tag{1}$$

where E_s is the energy contained in each symbol interval T_s, $\mathbf{u}_1^k=[u_1,u_2,\cdots,u_k]$ is the $\mathcal{M}$-ary information data sequence with elements chosen independently and with equal probability from the alphabet $\mathcal{U}=\{\pm1,\pm3,\cdots,\pm(M-1)\}$, and $\varphi(t,\mathbf{u}_1^k)$ is the excess phase carrying the information. The format of the excess phase is $\varphi(t,\mathbf{u}_1^k)=2\pi h_{\rm idx}\sum_{i=1}^{k}u_i g(t-iT_s+T_s)$, where $h_{\rm idx}=q/p$ is the rational modulation index, q and p are relatively prime, and $g(t)$ is the phase shaping pulse that is related to the frequency shaping pulse $f(t)$ as $g(t)=\int_0^t f(t)dt$.

[1]Some abbreviations might be redefined in each chapter for clarity of reading.

By convention, $g(t)$ is causal and its final value is normalized to $1/2$ after L symbol intervals. The phase shaping pulse with $L=1$ is called a full-response pulse, while those with $L>1$ are partial-response pulses.

The excess phase can be expressed as

$$\varphi(t,\mathbf{u}_1^k)=\pi h_{\text{idx}}\sum_{i=1}^{k-L}u_i+2\pi h_{\text{idx}}\sum_{i=k-L+1}^{k}u_i\,g(t-iT_s+T_s)=\varphi_k+\varphi_L(t,u_{k-L+1},u_{k-L+2},\cdots,u_k). \quad (2)$$

From (2), it is obvious that the modulated signal over the k-th interval can be specified once the phase state $\varphi_k=\pi h_{\text{idx}}\sum_{i=1}^{k-L}u_i$ and the most recent L symbols, i.e., $u_{k-L+1},\cdots,u_k$, are known. Therefore, the nature of the phase evolution can be described by a finite-state machine, where each state is represented by an L-dimensional vector $\boldsymbol{\sigma}_k=[\varphi_k,u_{k-L+1},\cdots,u_{k-1}]$. From this viewpoint, the CPM modulator can be viewed as a recursive convolutional encoder followed by a memoryless phase modulator [4].

2.1.1 Tilted-Phase Representation

The number of possible values for φ_k is $2p$ when q is odd, and p when q is even. When p is odd, φ_k can assume p possible values in the k-th interval, while the remaining p values are active in the $k+1$-st interval. Hence, the CPM modulator has a time-variant and periodic state transition diagram with a period of $2T_s$. To simplify such a structure, the *tilted-phase* representation of CPM signals was proposed in [87], where the resulting state description is time-invariant regardless of whether p is even or odd.

Under the tilted-phase representation, the CPM complex envelope becomes

$$x'(t,\mathbf{u}_1^k)=x(t,\mathbf{u}_1^k)\ \exp\left\{j\left(\pi h_{\text{idx}}(M-1)t/T_s\right)\right\}, \quad (3)$$

and the carrier frequency is shifted from f_c to $f_c-h_{\text{idx}}(M-1)/(2T_s)$. After some algebra the tilted excess phase can be rewritten as

$$\varphi'(t,\mathbf{u}_1^k)\ =\ \pi h_{\text{idx}}\sum_{i=1}^{k-L}u_i'+2\pi h_{\text{idx}}\sum_{i=k-L+1}^{k}u_i'\,g(t-iT_s+T_s)+\pi h_{\text{idx}}C(t-kT_s+T_s), \quad (4)$$

where $u_i'=u_i+M-1$, $u_i'\in\mathcal{U}'=\{0,2,\cdots,2(M-1)\}$, and

$$C(t)=(M-1)\frac{t}{T}+(M-1)(L-1)-2(M-1)\sum_{i=0}^{L-1}g(t+iT_s),\quad 0\leqslant t\leqslant T_s \quad (5)$$

is an adjusting term independent of the information sequence. The tilted phase state $\varphi'_k = \pi h_{\text{idx}} \sum_{i=1}^{k-L} u'_i$ has p possible values for p either even or odd. Consequently, under the tilted-phase representation, the total number of transition states within a CPM encoder is pM^{L-1}, and there are M branches starting from and merging into each state.

2.1.2 Laurent's Expansion

Inherent nonlinearity of CPM signals causes tough challenges in analyzing signal characteristics, such as autocorrelation and power spectrum densities, and designing practical demodulators regarding to the issues of implementation complexity and synchronization. In 1986, Laurent found that any binary CPM signal can be decomposed as the sum of a few linearly modulated pulses [25]. Later, it was observed by Mengali and Morelli that the same-type expansion was also valid for $\mathcal{M}$-ary CPM signals [88]. With aid of this new interpretation, many methodologies and tools developed for studying and devising linear modulation systems can be applied to CPM schemes with proper modifications.

Under the Laurent expansion, a binary CPM signal can be expressed by superposition of 2^{L-1} pulse-amplitude-modulated (PAM) waveforms as,

$$x(t, \mathbf{u}_1^k) = \sum_{i=1}^{k} \sum_{m=1}^{2^{L-1}} \alpha_{i,m} c_m(t - iT_s + T_s), \tag{6}$$

where

$$c_m(t) = c_0(t) \prod_{l=1}^{L-1} c_0(t + (l + L\beta_{m,l})T_s), \tag{7a}$$

$$\alpha_{i,m} = \exp\{j\pi h_{\text{idx}} \sum_{l=1}^{k} u_l - j\pi h_{\text{idx}} \sum_{l=1}^{L-1} u_{i-l}\beta_{m,l}\}, \tag{7b}$$

the basic pulse $c_0(t)$ is nonzero over $[0, 2T_s)$ and given by

$$c_0(t) = \begin{cases} \sin(2\pi h_{\text{idx}} g(t))/\sin(\pi h_{\text{idx}}), & 0 \leqslant t < LT_s, \\ \sin(\pi h_{\text{idx}} - 2\pi h_{\text{idx}} g(t - LT_s))/\sin(\pi h_{\text{idx}}), & LT_s \leqslant t < 2LT_s, \end{cases} \tag{7c}$$

and $\beta_{m,l}$ is the l-th digit of the binary representation of the index m. For example, the Laurent expansion of an MSK signal only has one component $c_1(t) = \sin(\pi t/(2T_s))$, which

is a partial-response waveform and valid over $[0, 2T_s)$. The Laurent expansion of an $\mathcal{M}$-ary CPM signal is more involved. The number of required PAM pulses is $(M-1)M^{L-1}$. Note that CPM signal energy is unevenly distributed among all pulses, and the sum of few of them usually conveys the most part of the signal energy.

2.1.3 Exponential Expansion

At the CPM receiver, a bank of matched filters is typically used to generated sufficient statistics for optimal coherent detection, called front-end processing. For the CPM signals with either the classical expression in (1) or the *tilted-phase* representation in (3), the number of required matched filters is M^L. With the Laurent expansion, this number reduces to $(M-1)M^{L-1}$ and a further decrease is possible by approximating CPM signals by the few most significant PAM waveforms. However, such waveforms are usually of complicated and partial-response format, with which the corresponding signal reception is not easy.

In light of the Fourier series, the CPM signals can be approximately represented by using a series of complex exponential functions as follows:

$$x'(t, \mathbf{u}_1^k) \approx \sum_{i=1}^{N_f} a_{k,i} \exp\left\{j2\pi f_i(t - kT_s + T_s)\right\}, \quad (k-1)T_s \leqslant t \leqslant kT_s, \tag{8}$$

where we consider the tilted-phase CPM signal without loss of generality, N_f is the number of basis functions (assumed to be odd) and f_i is the frequency of the i-th pulse with the complex coefficient $a_{k,i}$. The vector $\boldsymbol{a}_k = [a_{k,1}, a_{k,2}, \cdots, a_{k,N_f}]^T$ can be computed as $\boldsymbol{a}_k = \{\boldsymbol{R}_\rho^T\}^{-1}\boldsymbol{z}_k$, where $\boldsymbol{R}_\rho$ is the pulse-correlation matrix with the entry at the i-th row and l-th column given by

$$\{\boldsymbol{R}_\rho\}_{i,l} = \rho_{i,l} = \int_0^{T_s} \exp\{j2\pi(f_i - f_l)t\}dt, \tag{9a}$$

and $\boldsymbol{z}_k = [z_{k,1}, z_{k,N_2}, \cdots, z_{k,N_f}]^T$ is the projection of the CPM signal onto the basis functions such that

$$z_{k,i} = \int_{(k-1)T_s}^{kT_s} x'(t, \mathbf{u}_1^k) \exp\{-j2\pi f_i(t - kT_s + T_s)\}dt. \tag{9b}$$

Setting $f_i = (i - \lceil N_f/2 \rceil)/T_s$ in (8) yields an orthogonal exponential expansion (OEE), where $\lceil \cdot \rceil$ represents the ceil function.

Table 1: Performance of different CPM signal exponential expansions.

Exp. Expansions $\Longrightarrow$		OEE	OEE	OEE	SnOEE[a]	SnOEE[a]	nSnOEE[b]	nSnOEE[b]
4-3RC,	NPulses[c]	5	21	51	3	5	3	5
$h=\frac{1}{2}$	NMSE[c]	4.5×10^{-2}	9.7×10^{-3}	4.0×10^{-3}	9.4×10^{-2}	1.4×10^{-3}	5.9×10^{-3}	5.3×10^{-5}
GMSK,	NPulses[c]	5	21	51	3	5	3	5
$BT_s=0.3$	NMSE[c]	4.1×10^{-2}	9.7×10^{-3}	4.0×10^{-3}	7.3×10^{-4}	1.2×10^{-6}	1.8×10^{-4}	1.5×10^{-8}

[a] For SnOEE ν_{opt} is optimized by a coarse exhaustive search with grid spacing $\triangle\nu = 0.05$.

[b] For nSnOEE $\boldsymbol{f}_{opt}$ is estimated by a coarse exhaustive search with grid spacing $\triangle f_i = 0.05$.

[c] NPulses and NMSE stand for number of pulses and normalized MSE, respectively.

To achieve sufficient accuracy, the OEE usually requires a large number of basis functions. Huber and Liu [26] significantly reduced the required number of basis functions by using a symmetric nonorthogonal exponential expansion (SnOEE), i.e., setting $f_i = \nu(i - \lceil N_f/2 \rceil)/T_s$ with $0 < \nu < 1$. Note that the frequency spacing parameter ν must be optimized for each CPM scheme to minimize the mean-square-error between the actual and approximate signals. The main disadvantage of SnOEE is that the approximation is not sufficiently accurate for small values of N_f, even when the optimal ν_{opt} is selected.

To obtain a sufficiently accurate CPM signal representation with fewer basis functions, we propose an alternative approach, called a nonsymmetric nonorthogonal exponential expansion (nSnOEE), where the symmetric constraint in SnOEE, i.e., equal frequency spacing, is removed. Since the CPM signal power is typically concentrated in the frequency interval $[-1/T_s, 1/T_s]$, it is reasonable to assume $-1/T_s \leqslant f_i \leqslant 1/T_s$ for $1 \leqslant i \leqslant N_f$ in nSnOEE. The optimal frequency set $\boldsymbol{f}_{opt} = [f_{1,opt}, f_{2,opt}, \cdots, f_{N_f,opt}]$ is determined by the following criterion,

$$\boldsymbol{f}_{opt} = \arg \min_{|f_i| \leqslant 1} \mathbb{E} \left\{ \int_{(k-1)T_s}^{kT_s} \left| x'(t, \mathbf{u}_1^k) - \sum_{i=1}^{N_f} a_{k,i} \exp\{j2\pi f_i (t - kT_s + T_s)\} \right|^2 dt \right\}. \quad (10)$$

Solving the above non-convex optimization problem requires an exhaustive search for each candidate CPM scheme.

Table 1 compares the performance of OEE, SnOEE and nSnOEE in terms of the required

number of basis functions and the approximation accuracy. Two typical CPM schemes are considered, including GMSK with a 3 dB bandwidth $BT_s = 0.3$, and 4-ary, 3RC ($L = 3$ raised cosine) CPM with $h = 1/2$. Table 1 shows that OEE requires at least 21 orthogonal pulses to keep the normalized MSE on order of 10^{-3}. The nonorthogonal exponential expansions (nOEE) with suboptimal parameter sets can achieve a better approximation accuracy with only 3 pulses. The required suboptimal parameter sets can be estimated by using coarse exhaustive searches with the grid spacing $\triangle\nu = 0.05$ for SnOEE, or $\triangle f_i = 0.05/T_s$ for nSnOEE. Observe that nSnOEE provides a more accurate CPM signal representation than SnOEE, since nSnOEE has fewer constraints.

2.2 Joint Timing and Phase Recovery

The received CPM signal with unknown time shift and carrier phase offset is given by

$$y(t) = \exp\{j\phi_k\}x(t-\tau_k, \mathbf{u}_1^k) + n(t), \;\; (k-1)T_s - \tau_k \leqslant t < kT_s - \tau_k \tag{11}$$

where we assume both time shift τ_k and phase shift ϕ_k are fixed over each symbol interval, and $n(t)$ is the additive white Gaussian noise (AWGN) with two-sided power spectral density $N_o/2$. Using exponential expansions the received CPM signal can be expressed as

$$y(t) = \sum_{i=1}^{N_f} a_{k,i} \exp\{j(\phi_k - 2\pi f_i \tau_k)\} \exp\{j2\pi f_i(t - kT_s + T_s)\} + n(t) + \zeta(t), \tag{12}$$

where $\zeta(t)$ is the residual error caused by the signal approximation. Let the front-end filters be matched to the exponential basis waveforms. The output of the l-th matched filter, sampled at the instant of τ_k' and compensated with the phase offset ϕ_k', is given by

$$\begin{aligned} y_{k,l} &= y(t) \otimes \exp\{-j(\phi_k' + 2\pi f_l(T_s - t))\}|_{t=kT_s+\tau_k'} \\ &= \sum_{i=1}^{N_f} a_{k,i}\rho_{i,l} \exp\{j(\phi_k - \phi_k' - 2\pi f_i(\tau_k - \tau_k'))\} + n_{k,l} + \zeta_{k,l} + \xi_{k,l}, \end{aligned} \tag{13}$$

where $n_{k,l}$ is a zero-mean Gaussian random variable with variance N_o, given by

$$n_{k,l} = \int_{(k-1)T_s-\tau_k'}^{kT_s-\tau_k'} n(t) \exp\{-j(\phi_k' + 2\pi f_l(t - kT_s + T_s))\}dt, \tag{14a}$$

the residual $\zeta_{k,l}$ is

$$\zeta_{k,l} = \int_{(k-1)T_s-\tau'_k}^{kT_s-\tau'_k} \zeta(t) \exp\{-j(\phi'_k + 2\pi f_l(t - kT_s + T_s))\}dt, \tag{14b}$$

and $\xi_{k,l}$ is the intersymbol interference (ISI) resulted from imperfect timing and can be expressed as

$$\xi_{k,l} = \sum_{i=1}^{N_f} a_{k,i} \exp\{j(\phi_k - \phi'_k - 2\pi f_i(\tau_k - \tau'_k))\} \int_{(k-1)T_s+\tau'_k}^{(k-1)T_s+\tau_k} \exp\{j2\pi(f_i - f_l)t\}dt, \tag{15}$$

by assuming $\tau_k > \tau'_k$. Note that $\xi_{k,l}$ vanishes when $\tau_k = \tau'_k$. Without loss of generality, we set $l = 1$ hereafter and drop the index l for simplicity. After some manipulations, (13) can be rewritten in the matrix form

$$y_k = \boldsymbol{s}_k \boldsymbol{\theta}_k + v_k, \tag{16}$$

where the vector $\boldsymbol{s}_k = [s_{k,1}, s_{k,2}, \cdots, s_{k,N_f}]$ with $s_{k,i} = a_{k,i}\rho_{i,1} \exp\{-j(\phi'_k - 2\pi f_i \tau'_k)\}$ carries both data and synchronization information, the state vector $\boldsymbol{\theta}_k = [\theta_{k,1}, \theta_{k,2}, \cdots, \theta_{k,N_f}]^T$ with $\theta_{k,i} = \exp\{j(\phi_k - 2\pi f_i \tau_k)\}$ conveys the complete information of the time shift and phase offset, and $v_k = n_k + \zeta_k + \xi_k$. The statistics of v_k are difficult, if not impossible, to be determined in practice because of the existence of ζ_k and ξ_k. However, because of the randomness of transmitting signals and sampling instants, it is reasonable from both analysis and practical aspects to assume that v_k is a white noise with zero mean and variance r_v in the estimator design, although the real model of v_k has to be considered in simulations. The effect of this modeling error will be discussed in more detail in the section 2.3. Furthermore, when the considered exponential expansion is sufficiently accurate and the CPM signal is approximately synchronized, ζ_k and ξ_k become negligible.

2.2.1 Robust Time and Phase Recovery on Time-invariant Channels

On time-invariant channels the state vector $\boldsymbol{\theta}_k$ is fixed, i.e., $\boldsymbol{\theta}_k = \boldsymbol{\theta}$ or $[\phi_k, \tau_k] = [\phi, \tau]$. Moreover, we assume that the time shift and phase offset are independently and uniformly distributed over $[-T_s/2, T_s/2]$ and $[-\pi, \pi]$, respectively. Therefore, the state vector $\boldsymbol{\theta}$ has zero-mean and its covariance matrix $\boldsymbol{R}_{\boldsymbol{\theta}}$ is

$$\{\boldsymbol{R}_{\boldsymbol{\theta}}\}_{i,j} = \mathbb{E}\{\theta_i \theta_j^*\} = \mathrm{sinc}(\pi(f_j - f_i)T_s), \tag{17}$$

where $\mathrm{sinc}(x) = \sin(x)/x$ and $(\cdot)^*$ denotes conjugate transpose.

The received samples $\boldsymbol{y}_k = [y_1, y_2, \cdots, y_k]^T$ can be expressed as

$$\boldsymbol{z}_k = \boldsymbol{S}_k\boldsymbol{\theta} + \boldsymbol{v}_k, \tag{18}$$

where $\boldsymbol{S}_k = [\boldsymbol{s}_1^T, \boldsymbol{s}_2^T, \cdots, \boldsymbol{s}_k^T]^T$ is a $k \times N_f$ matrix, and $\boldsymbol{v}_k = [v_1, v_2, \cdots, v_k]^T$ is a k-dimensional noise vector with covariance $r_v\boldsymbol{I}_k$, and $\boldsymbol{I}_k$ is a $k \times k$ identity matrix. The MMSE estimate of $\boldsymbol{\theta}$ on the k-th time interval is [89]

$$\hat{\boldsymbol{\theta}}_k = \boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{S}_k^* \left(\boldsymbol{S}_k\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{S}_k^* + r_v\boldsymbol{I}_k\right)^{-1} \boldsymbol{z}_k, \tag{19}$$

and its error covariance matrix $\boldsymbol{M}_k$ is

$$\boldsymbol{M}_k = \mathrm{Cov}(\hat{\boldsymbol{\theta}}_k) = \boldsymbol{R}_{\boldsymbol{\theta}} - \boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{S}_k^* \left(\boldsymbol{S}_k\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{S}_k^* + r_v\boldsymbol{I}_k\right)^{-1} \boldsymbol{S}_k\boldsymbol{R}_{\boldsymbol{\theta}}. \tag{20}$$

If reliable timing and phase information at the k-th epoch can be extracted from $\hat{\boldsymbol{\theta}}_k$ as $\hat{\tau}_k$ and $\hat{\phi}_k$, respectively, the sampling instant τ'_{k+1} and the phase compensation ϕ'_{k+1} at the $k+1$-st epoch are adaptively set to $\hat{\tau}_k$ and $\hat{\phi}_k$, respectively.

Let $\boldsymbol{\Phi}_k = [\angle\hat{\theta}_{k,1}, \angle\hat{\theta}_{k,2}, \cdots, \angle\hat{\theta}_{k,N_f}]^T$, where $\angle(\cdot)$ returns the phase of a complex variable on the interval $[-\pi, \pi]$. If the unknown time shift and phase offset are sufficiently small such that $\phi_k - 2\pi f_i\tau_k$ lies on $[-\pi, \pi]$ for $1 \leqslant i \leqslant N_f$, then phase wrapping rarely occurs when computing $\boldsymbol{\Phi}_k$. As a result, the estimates of τ_k and ϕ_k can be expressed as

$$\boldsymbol{\Phi}_k = \boldsymbol{G}\ [\hat{\phi}_k\ \ \hat{\tau}_k]^T, \tag{21}$$

where $\boldsymbol{G} = \left[[1, -2\pi f_1]^T, [1, -2\pi f_2]^T, \cdots, [1, -2\pi f_{N_f}]^T\right]^T$. The time shift and phase offset at the k-th epoch can be estimated as

$$[\hat{\phi}_k\ \ \hat{\tau}_k]^T = (\boldsymbol{G}^T\boldsymbol{G})^{-1}\boldsymbol{G}^T\boldsymbol{\Phi}_k. \tag{22}$$

When the time shift or phase offset is large, $\phi_k - 2\pi f_i\tau_k$ might fall outside the interval $[-\pi, \pi]$ and must be wrapped around into $[-\pi, \pi]$ by introducing $\pm 2\pi$ phase jumps. Hence, (21) is no longer valid. Instead of estimating $\boldsymbol{\theta}$ directly in (19), we propose to estimate $\boldsymbol{\epsilon}_k = [\epsilon_{k,1}, \cdots, \epsilon_{k,N_f}] = \boldsymbol{\Lambda}_P(\hat{\boldsymbol{\theta}}^*_{k-1})\boldsymbol{\theta}$, at the k-th time epoch , where $\epsilon_{k,i} = \exp\{j(\phi - \hat{\phi}_{k-1} - 2\pi f_i(\tau - \hat{\tau}_{k-1}))\}$ and $\boldsymbol{\Lambda}_P([c_1, c_2, \cdots, c_{N_c}]^T)$ represents an $N_c \times N_c$ diagonal matrix with

$\exp\{j\angle c_i\}$ as the i-th diagonal element. Denote the estimate of $\boldsymbol{\epsilon}_k$ as $\hat{\boldsymbol{\epsilon}}_k = [\hat{\epsilon}_{k,1}, \cdots, \hat{\epsilon}_{k,N_f}]$ with $\epsilon_{k,i} = \exp\{j(\hat{\phi}_k - \hat{\phi}_{k-1} - 2\pi f_i(\hat{\tau}_k - \hat{\tau}_{k-1}))\}$. Provided $\hat{\boldsymbol{\theta}}_{k-1}$ is close to $\boldsymbol{\theta}$, the phase of $\epsilon_{k,i}$ is small and, thus, no phase wrapping is necessary when computing $\angle(\hat{\epsilon}_{k,i})$, i.e., $\angle(\hat{\theta}_{k,i}) - \angle(\hat{\theta}_{k-1,i})$, even when the unknown time shift or phase offset is large. Letting $\boldsymbol{F}_k = \boldsymbol{\Lambda}_P(\hat{\boldsymbol{\epsilon}}_k^*)$ and $\boldsymbol{H}_k = \boldsymbol{s}_k\boldsymbol{\Lambda}_P(\hat{\boldsymbol{\theta}}_{k-1})$, we can express (16) in an alternative form by employing a time-variant state-space model as follows:

$$\begin{aligned}
\boldsymbol{\epsilon}_{k+1} &= \boldsymbol{\Lambda}_P(\hat{\boldsymbol{\theta}}_k^*)\boldsymbol{\theta}, \\
&= \boldsymbol{\Lambda}_P(\hat{\boldsymbol{\theta}}_k^*)\boldsymbol{\Lambda}_P(\hat{\boldsymbol{\theta}}_{k-1})\boldsymbol{\Lambda}_P(\hat{\boldsymbol{\theta}}_{k-1}^*)\boldsymbol{\theta}, \\
&= \boldsymbol{\Lambda}_P(\hat{\boldsymbol{\epsilon}}_k^*)\boldsymbol{\Lambda}_P(\hat{\boldsymbol{\theta}}_{k-1}^*)\boldsymbol{\theta}, \\
&= \boldsymbol{F}_k\boldsymbol{\epsilon}_k, \qquad (23a) \\
y_k &= \boldsymbol{s}_k\boldsymbol{\theta}_k + v_k, \\
&= \boldsymbol{s}_k\boldsymbol{\Lambda}_P(\hat{\boldsymbol{\theta}}_{k-1})\boldsymbol{\epsilon}_k + v_k, \\
&= \boldsymbol{H}_k\boldsymbol{\epsilon}_k + v_k. \qquad (23b)
\end{aligned}$$

With Kalman Filtering [90], the filtered estimates of $\boldsymbol{\theta}_k$ and $\boldsymbol{\epsilon}_k$ are updated as follows:

$$\begin{aligned}
\hat{\boldsymbol{\epsilon}}_{k+1} &= \boldsymbol{F}_k\hat{\boldsymbol{\epsilon}}_k + \boldsymbol{K}_{f,k+1}(z_{k+1} - \boldsymbol{H}_{k+1}\boldsymbol{F}_k\hat{\boldsymbol{\epsilon}}_k) \qquad (24a) \\
\hat{\boldsymbol{\theta}}_{k+1} &= \Lambda_P\left(\hat{\boldsymbol{\epsilon}}_{k+1}\right)\hat{\boldsymbol{\theta}}_k, \qquad (24b) \\
\boldsymbol{K}_{f,k+1} &= \boldsymbol{F}_k\boldsymbol{P}_k\boldsymbol{F}_k^*\boldsymbol{H}_{k+1}^*/r_{e,k+1}, \qquad (24c) \\
r_{e,k+1} &= r_v + \boldsymbol{H}_{k+1}\boldsymbol{F}_k\boldsymbol{P}_k\boldsymbol{F}_k^*\boldsymbol{H}_{k+1}^*, \qquad (24d) \\
\mathrm{Cov}(\hat{\boldsymbol{\epsilon}}_{k+1}) &= \boldsymbol{P}_{k+1} = \boldsymbol{F}_k\boldsymbol{P}_k\boldsymbol{F}_k^* - r_{e,k+1}\boldsymbol{K}_{f,k+1}\boldsymbol{K}_{f,k+1}^*, \qquad (24e)
\end{aligned}$$

where $\boldsymbol{K}_{f,k}$ is the Kalman filter gain and $\boldsymbol{P}_k$ is the error covariance matrix. The filter is initialized at $\hat{\boldsymbol{\theta}}_0 = [1, 1, \cdots, 1]$ by assuming $\hat{\phi}_0 = 0$ and $\hat{\tau}_0 = 0$, $\hat{\boldsymbol{\epsilon}}_1 = \boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{H}_1^* z_1/r_{e,1}$, $\boldsymbol{P}_1 = \boldsymbol{R}_{\boldsymbol{\theta}} - \boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{H}_1^*\boldsymbol{H}_1\boldsymbol{R}_{\boldsymbol{\theta}}/r_{e,1}$, and $r_{e,1} = r_v + \boldsymbol{H}_1\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{H}_1^*$. Let $\triangle\boldsymbol{\Phi}_k \stackrel{\Delta}{=} [\angle\hat{\epsilon}_{k,1}, \angle\hat{\epsilon}_{k,1}, \cdots, \angle\hat{\epsilon}_{k,N_f}]^T$. Similar to (21), one can update the time and phase estimates as,

$$[\hat{\phi}_k \;\; \hat{\tau}_k]^T = [\hat{\phi}_{k-1} \;\; \hat{\tau}_{k-1}]^T + \beta_k(\boldsymbol{G}^T\boldsymbol{G})^{-1}\boldsymbol{G}^T\triangle\boldsymbol{\Phi}_k, \qquad (25)$$

where the step size β_k is set to 1, but one could also assume other values in order to adjust the convergence rate of the proposed estimator.

2.2.2 Robust Time and Phase Recovery on Time-variant Channels

We model time-variant channels as a standard Wiener process, i.e.,

$$\tau_{k+1} = \tau_k + w_{\tau,k}, \quad \text{and}, \quad \phi_{k+1} = \phi_k + w_{\phi,k}, \tag{26}$$

where the initial time shift and phase offset, τ_0 and ϕ_0, lie in $[-T_s/2, T_s/2]$ and $[-\pi, \pi]$, respectively as before, and $w_{\tau,k}$ and $w_{\phi,k}$ are the k-th timing and phase jitters, respectively. Moreover, we assume that the parameter jitters are independent of the channel noise. Letting $\boldsymbol{\gamma}_k \stackrel{\Delta}{=} [\gamma_{k,1}, \gamma_{k,2}, \cdots, \gamma_{k,N_f}]^T$, where $\gamma_{k,i} = \exp\{j(w_{\phi,k} - 2\pi f_i w_{\tau,k})\}$, the state transition in (23a) is revised to fit time-variant channels as

$$\begin{aligned}
\boldsymbol{\epsilon}_{k+1} &= \boldsymbol{\Lambda}_P(\hat{\boldsymbol{\theta}}_k^*)\boldsymbol{\theta}_{k+1}, \\
&= \boldsymbol{\Lambda}_P(\hat{\boldsymbol{\theta}}_k^*)\boldsymbol{\Lambda}_P(\hat{\boldsymbol{\theta}}_{k-1})\boldsymbol{\Lambda}_P(\hat{\boldsymbol{\theta}}_{k-1}^*)\boldsymbol{\Lambda}_P(\boldsymbol{\gamma}_k)\boldsymbol{\theta}_k, \\
&= \boldsymbol{\Lambda}_P(\hat{\boldsymbol{\epsilon}}_k^*)\boldsymbol{\Lambda}_P(\boldsymbol{\gamma}_k)\boldsymbol{\Lambda}_P(\hat{\boldsymbol{\theta}}_{k-1}^*)\boldsymbol{\theta}_k, \\
&= \boldsymbol{F}_k\boldsymbol{\Lambda}_P(\boldsymbol{\gamma}_k)\boldsymbol{\epsilon}_k,
\end{aligned} \tag{27}$$

which falls in a class of state-space models with multiplicative noise. Pursuing the optimal solution of this problem is generally difficult. However, when the time and phase jitters are small at each symbol interval, a good approximation of (27) can be obtained by employing a Taylor expansion on $\boldsymbol{\gamma}_k$ and assuming $\hat{\boldsymbol{\epsilon}}_k \approx \boldsymbol{\epsilon}_k$,

$$\begin{aligned}
\boldsymbol{\epsilon}_{k+1} &\approx \boldsymbol{F}_k\Big(\boldsymbol{I}_{N_f} + \boldsymbol{\Lambda}_P(\boldsymbol{G}\boldsymbol{w}_k)\Big)\boldsymbol{\epsilon}_k, \\
&= \boldsymbol{F}_k\boldsymbol{\epsilon}_k + \boldsymbol{\Lambda}_P(\hat{\boldsymbol{\epsilon}}_k^*)\boldsymbol{\Lambda}_P(\boldsymbol{\epsilon}_k)\boldsymbol{G}\boldsymbol{w}_k, \\
&\approx \boldsymbol{F}_k\boldsymbol{\epsilon}_k + \boldsymbol{G}\boldsymbol{w}_k,
\end{aligned} \tag{28}$$

where $\boldsymbol{w}_k = [jw_{\phi,k} \;\; jw_{\tau,k}]^T$ has covariance matrix $\boldsymbol{Q}$. Under the assumption that $[\phi_k \;\; \tau_k]^T$ is uncorrelated with $\boldsymbol{w}_k$, $\boldsymbol{\epsilon}_k$ is thus uncorrelated with $\boldsymbol{w}_k$, which ensures the optimality of the Kalman filter in the MMSE sense. Therefore, for time-variant channels, the filtered estimates of $\boldsymbol{\theta}_k$ and $\boldsymbol{\epsilon}_k$ can be updated in a similar way as in (24a) and (24b), except that the updating recursions of the Kalman gain and the error covariance are revised as follows:

$$\boldsymbol{K}_{f,k+1} = (\boldsymbol{F}_k \boldsymbol{P}_k \boldsymbol{F}_k^* + \boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^*)\boldsymbol{H}_{k+1}^* / r_{e,k+1}, \tag{29a}$$

$$r_{e,k+1} = r_v + \boldsymbol{H}_{k+1}(\boldsymbol{F}_k \boldsymbol{P}_k \boldsymbol{F}_k^* + \boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^*)\boldsymbol{H}_{k+1}^*, \tag{29b}$$

$$\boldsymbol{P}_{k+1} = \boldsymbol{F}_k \boldsymbol{P}_k \boldsymbol{F}_k^* + \boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^* - r_{e,k+1}\boldsymbol{K}_{f,k+1}\boldsymbol{K}_{f,k+1}^*. \tag{29c}$$

Finally, one can obtain estimates of the time shift and phase offset through (25).

2.3 Convergence Analysis of The Proposed Synchronizer

The proposed synchronizer on either time-invariant or time-variant channels is essentially a Kalman filter derived from the set of time-variant state-space models given in (23) and (28). However, the asymptotic stability of a Kalman filter is generally not guaranteed. In this section, we first introduce a few known results on observability and stability of dynamic linear systems, and then utilize them to analyze the convergence of the proposed synchronizer. The effect of statistical modeling errors on the convergence is also discussed.

2.3.1 Asymptotic Behavior of The Proposed Synchronizer

The concept of uniform observability and a well-known result on the asymptotic behavior of time-variant linear systems are summarized below. More details are available in [91] and [92].

Definition 1: The sequence of the matrix pairs $\{[\boldsymbol{F}_k, \boldsymbol{H}_k]\}_{k=1}^{+\infty}$ is *uniformly observable* if for every $k > 0$ there exists a positive integer m_k, such that the observability Gramian matrix $\boldsymbol{O}_k$ is positive definite, i.e.,

$$\boldsymbol{O}_k = \sum_{i=k}^{k+m_k-1} \boldsymbol{T}_{\boldsymbol{F}}^*(i,k)\boldsymbol{H}_i^*\boldsymbol{H}_i\boldsymbol{T}_{\boldsymbol{F}}(i,k) > \boldsymbol{0}, \tag{30}$$

where $\boldsymbol{T}_{\boldsymbol{F}}(i,k)$ is the state transition matrix associated with the sequence of matrices $\{\boldsymbol{F}_k\}_{k=1}^{+\infty}$ and defined as $\boldsymbol{T}_{\boldsymbol{F}}(k,k) = \boldsymbol{I}$ and $\boldsymbol{T}_{\boldsymbol{F}}(i,k) = \boldsymbol{F}_{i-1}\boldsymbol{T}_{\boldsymbol{F}}(i-1,k)$. If $m_k = m$ for all k, the sequence $\{[\boldsymbol{F}_k, \boldsymbol{H}_k]\}_{k=1}^{+\infty}$ is *uniformly m-step observable*. There are a few remarks on *Definition* 1:

1. The positive definiteness of $\boldsymbol{O}_k$ intuitively implies that if the trajectory of the state vector does not decay (or converge to the origin), then the output signal can be measured or observed.

2. Conversely, if the output signals decay to zero, the state variables are unobservable. The associated Kalman filter is unable to accurately estimate unknown states.

Proposition 1: For a standard finite-dimensional state space model, if $\{[\boldsymbol{F}_k, \boldsymbol{H}_k]\}_{k=1}^{+\infty}$ is *uniformly observable* and all parameter matrices, including $\boldsymbol{F}_k$, $\boldsymbol{H}_k, \boldsymbol{G}$ and $\boldsymbol{Q}$, are bounded, then the Kalman filtered error covariance matrix $\boldsymbol{P}_k$ is also bounded $\forall\ k$. In addition, if the process noise $\boldsymbol{w}_k$ is absent, i.e, $\boldsymbol{w}_k \equiv \boldsymbol{0}\ \forall\ k$, then $\boldsymbol{P}_k$ converges exponentially to zero matrix as $k \to +\infty$.

Recall that for the proposed synchronizer, $\boldsymbol{F}_k = \boldsymbol{\Lambda}_P(\hat{\boldsymbol{\theta}}_{k-1}\hat{\boldsymbol{\theta}}_k^*)$ and $\boldsymbol{H}_k = \boldsymbol{s}_k\boldsymbol{\Lambda}_P(\hat{\boldsymbol{\theta}}_{k-1})$. The vector $\boldsymbol{s}_k = \tilde{\boldsymbol{s}}_k\Lambda_P(\boldsymbol{\theta}_k'^*)$, where $\boldsymbol{\theta}_k' = [\theta_{k,1}', \theta_{k,2}', \cdots, \theta_{k,N_f}']$ with $\theta_{k,i}' = \exp\{j(\phi_k' - 2\pi f_i\tau_k')\}$, conveys the synchronization information, and $\tilde{\boldsymbol{s}}_k$ carries the data information at the k-th symbol interval. Before showing the uniform observability of the proposed estimator, we prove the following lemma:

Lemma 1: For the proposed synchronizer defined by (23) and (28), the sequence $\{[\boldsymbol{F}_k, \boldsymbol{H}_k]\}_{k=1}^{+\infty}$ is *uniformly N_f-step observable* with probability 1 if we set $\boldsymbol{\theta}_k' = \hat{\boldsymbol{\theta}}_{k-1}$.

Proof: Substituting $\boldsymbol{F}_k$ and $\boldsymbol{H}_k$ in (23) into (30) and letting $m_k = N_f$, one can obtain

$$\boldsymbol{O}_k\Lambda_P(\hat{\boldsymbol{\theta}}_{k-1}^*)\left\{\sum_{i=k}^{k+N_f-1}\Lambda_P(\hat{\boldsymbol{\theta}}_{i-1})\tilde{\boldsymbol{s}}_i^*\tilde{\boldsymbol{s}}_i\Lambda_P(\hat{\boldsymbol{\theta}}_{i-1}^*)\right\}\Lambda_P(\hat{\boldsymbol{\theta}}_{k-1}). \tag{31}$$

Constructing the matrix $\boldsymbol{B}_k = [\Lambda_P(\hat{\boldsymbol{\theta}}_{k-1})\tilde{\boldsymbol{s}}_k^*, \cdots, \Lambda_P(\hat{\boldsymbol{\theta}}_{k+N_f-2})\tilde{\boldsymbol{s}}_{k+N_f-1}^*]$ with dimension $N_f \times N_f$, one has $\boldsymbol{O}_k = \Lambda_P(\hat{\boldsymbol{\theta}}_{k-1}^*)\boldsymbol{B}_k\boldsymbol{B}_k^*\Lambda_P(\hat{\boldsymbol{\theta}}_{k-1})$. Since $\Lambda_P(\hat{\boldsymbol{\theta}}_{k-1})$ is a unitary matrix, the sufficient and necessary condition for the positive definiteness of $\boldsymbol{O}_k$ is that $\boldsymbol{B}_k$ has full rank. Given $\boldsymbol{\theta}_k' = \hat{\boldsymbol{\theta}}_{k-1}$, $\boldsymbol{B}_k$ is thus a random matrix even when the data sequence, $\{\tilde{\boldsymbol{s}}\}_{i=k}^{k+N_f-1}$, is known. Such randomness is caused by the feedback of the parameter estimates. Now suppose the rank of $\boldsymbol{B}_k$ is l which is less than N_f. Without loss of generality, we further assume that the first l column vectors in $\boldsymbol{B}_k$ are linearly independent. Hence, there exist

constants $[\alpha_1, \alpha_2, \cdots, \alpha_l] \neq 0$, such that

$$\alpha_1 \Lambda_P(\hat{\boldsymbol{\theta}}^*_{k-1})\tilde{\boldsymbol{s}}^*_k + \cdots + \alpha_l \Lambda_P(\hat{\boldsymbol{\theta}}^*_{k+l-2})\tilde{\boldsymbol{s}}^*_{k+l-1}\Lambda_P(\hat{\boldsymbol{\theta}}^*_{k+l-1})\tilde{\boldsymbol{s}}^*_{k+l}, \tag{32}$$

which can be rewritten as

$$\begin{aligned}\tilde{\alpha}_1 \Lambda_P(\hat{\boldsymbol{\theta}}^*_{k-1})\tilde{\boldsymbol{s}}^*_k \; + \; \cdots + \tilde{\alpha}_l \Lambda_P(\hat{\boldsymbol{\theta}}^*_{k+l-2})\tilde{\boldsymbol{s}}^*_{k+l-1} \\ = \; \Lambda_P([\exp\{j2\pi f_1 \hat{\tau}_{k+l-1}\}, \cdots, \exp\{j2\pi f_{N_f} \hat{\tau}_{k+l-1}\}]^T)\tilde{\boldsymbol{s}}^*_{k+l},\end{aligned} \tag{33}$$

where $\tilde{\alpha}_i = \alpha_i \exp\{j\hat{\phi}_{k+l-1}\}$. Given $\tilde{\boldsymbol{s}}_k, \cdots, \tilde{\boldsymbol{s}}_{k+l}$ and $\hat{\boldsymbol{\theta}}_{k-1}, \cdots, \hat{\boldsymbol{\theta}}_{k+l-2}$, the $l+1$ unknown coefficients, including $\tilde{\alpha}_1, \cdots, \tilde{\alpha}_l$ and $\hat{\tau}_{k+l-1}$, must be determined from the N_f nonlinear irreducible equations implied by (33). Specially, the solution for $\hat{\tau}_{k+l-1}$ either does not exist or could have a unique solution lying on $[-T_s/2, T_s/2]$. Because of randomness of the timing estimate, the probability that $\hat{\tau}_{k+l-1}$ satisfies (33) is infinitely small and so is the probability that $\boldsymbol{B}_k$ is singular. Consequently, the proposed synchronizer is *uniformly N_f-step observable* with probability 1.□

If we do not update the sampling instant τ'_k and phase compensation ϕ'_k at every symbol interval, then the result in *Lemma* 1 may not hold. For example, if one updates the timing and phase estimates every N_f symbol intervals, i.e, letting $\boldsymbol{\theta}'_k = \cdots = \boldsymbol{\theta}'_{k+N_f-1} = \hat{\boldsymbol{\theta}}_{k-1}$, then the positive definiteness of $\boldsymbol{O}_k$ is simply determined by the rank of $\tilde{\boldsymbol{B}}_k = [\tilde{\boldsymbol{s}}^*_k, \cdots, \tilde{\boldsymbol{s}}^*_{k+N_f-1}]$. In this case, if the dimension of the employed exponential expansion is greater than the dimension of the CPM signal space, i.e., $N_f > M^L$, then both $\tilde{\boldsymbol{B}}_k$ and $\boldsymbol{O}_k$ are singular and, thus, the proposed synchronizer is unobservable. Even if $N_f \leqslant M^L$, the uniform N_f-step observability of the proposed synchronizer depends on whether or not any N_f consecutive CPM signals are linearly independent. For data-aided synchronization, this requirement imposes a strict constraint on the design of training sequences; for decision-directed estimation, it is even worse since the randomly generated signals might impair the observability of the Kalman filter.

Now one can establish the stability of the proposed synchronizer as follows:

Theorem 1: For the proposed synchronizer with the adaptively updated sampling instants, the error variances of the timing and phase estimates are bounded on time-variant channels, and converge exponentially to zero on time-invariant channels.

Proof: The proof is immediate from *Proposition* 1 and *Lemma* 1. □

2.3.2 Effect of Modeling Errors on Convergence

For dynamic linear systems, optimal Kalman filtering requires perfect knowledge of the initial error covariance matrix and the covariance matrices of the process noise and measurement noise. However, in most practical applications, these covariances are known only approximately. The effect of statistical modeling errors on the stability of Kalman filtering has been studied extensively in literature. In this subsection, we will apply some known results to the proposed synchronizer. Moreover, a new result is presented in case of the absence of timing and phase jitters.

Consider the time-variant linear system defined in (23) and (28). Let $\bar{r}_v$, $\bar{\boldsymbol{Q}}$ and $\bar{\boldsymbol{P}}_1$ be the incorrect modeling parameters which are used to design the proposed synchronizer. The resulting Kalman gain, the variance of innovations and the error covariance are denoted as $\bar{\boldsymbol{K}}_{f,k}$, $\bar{r}_{e,k}$ and $\bar{\boldsymbol{P}}_k$, respectively, which are computed from (29). Substituting $\bar{\boldsymbol{K}}_{f,k}$ into (24a), the actual estimate of $\boldsymbol{\epsilon}_k$ is

$$\bar{\hat{\boldsymbol{\epsilon}}}_{k+1} = \boldsymbol{F}_k \bar{\hat{\boldsymbol{\epsilon}}}_k + \bar{\boldsymbol{K}}_{f,k+1}(z_{k+1} - \boldsymbol{H}_{k+1}\boldsymbol{F}_k \bar{\hat{\boldsymbol{\epsilon}}}_k). \tag{34}$$

Although $\bar{\hat{\boldsymbol{\epsilon}}}_{k+1}$ is no longer the linear MMSE estimate, it is still unbiased. After some algebra, the actual error covariance $\breve{\boldsymbol{P}}_k = \mathrm{Cov}(\bar{\hat{\boldsymbol{\epsilon}}}_k)$ can be updated as

$$\breve{\boldsymbol{P}}_{k+1} = \bar{\boldsymbol{A}}_{k+1}\boldsymbol{F}_k \breve{\boldsymbol{P}}_k \boldsymbol{F}_k^* \bar{\boldsymbol{A}}_{k+1}^* + r_v \bar{\boldsymbol{K}}_{f,k+1}\bar{\boldsymbol{K}}_{f,k+1}^* + \bar{\boldsymbol{A}}_{k+1}\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^* \bar{\boldsymbol{A}}_{k+1}^*, \tag{35}$$

with $\breve{\boldsymbol{P}}_1 = \bar{\boldsymbol{P}}_1$ and $\bar{\boldsymbol{A}}_{k+1} = \boldsymbol{I}_{N_f} - \bar{\boldsymbol{K}}_{f,k+1}\boldsymbol{H}_{k+1}$. Note that for $k \geqslant 2$ $\breve{\boldsymbol{P}}_k$ is not equal to $\bar{\boldsymbol{P}}_k$ which is only used to compute the designed Kalman gain and can be updated as

$$\bar{\boldsymbol{P}}_{k+1} = \bar{\boldsymbol{A}}_{k+1}\boldsymbol{F}_k \bar{\boldsymbol{P}}_k \boldsymbol{F}_k^* \bar{\boldsymbol{A}}_{k+1}^* + \bar{r}_v \bar{\boldsymbol{K}}_{f,k+1}\bar{\boldsymbol{K}}_{f,k+1}^* + \bar{\boldsymbol{A}}_{k+1}\boldsymbol{G}\bar{\boldsymbol{Q}}\boldsymbol{G}^* \bar{\boldsymbol{A}}_{k+1}^*. \tag{36}$$

A few observations on the actual error covariance $\breve{\boldsymbol{P}}_k$ are stated below:

1. It is shown in [93] that if $\{[\boldsymbol{F}_k, \boldsymbol{H}_k]\}_{k=1}^{+\infty}$ is *uniformly observable*, and $\bar{r}_v$, $\bar{\boldsymbol{Q}}$ and $\bar{\boldsymbol{P}}_1$ are bounded, then both $\breve{\boldsymbol{P}}_k$ and $\bar{\boldsymbol{P}}_k$ are bounded. This implies that, with statistical modeling errors, the proposed synchronizer converges with probability 1.

2. If $\bar{r}_v \leqslant r_v$, $\bar{\boldsymbol{Q}} \leqslant \boldsymbol{Q}$ and $\bar{\boldsymbol{P}}_1 \leqslant \boldsymbol{P}_1$, then $\bar{\boldsymbol{P}}_k \leqslant \boldsymbol{P}_k \leqslant \breve{\boldsymbol{P}}_k, \forall\, k$; if $\bar{r}_v \geqslant r_v$, $\bar{\boldsymbol{Q}} \geqslant \boldsymbol{Q}$ and $\bar{\boldsymbol{P}}_1 \geqslant \boldsymbol{P}_1$, then $\boldsymbol{P}_k \leqslant \breve{\boldsymbol{P}}_k \leqslant \bar{\boldsymbol{P}}_k, \forall\, k$ [94].

3. If $\bar{r}_v = \alpha r_v$, $\bar{\boldsymbol{Q}} = \alpha \boldsymbol{Q}$ and $\bar{\boldsymbol{P}}_1 = \alpha \boldsymbol{P}_1$, it is shown in [95] that the Kalman filter is completely insensitive to these scaling errors, i.e., $\bar{\boldsymbol{K}}_{f,k} = \boldsymbol{K}_{f,k}$ and thus $\boldsymbol{P}_k = \breve{\boldsymbol{P}}_k, \forall\, k$.

Recall that the initial error covariance for the proposed synchronizer is

$$\boldsymbol{P}_1 = \boldsymbol{R}_1 - \frac{1}{r_v + \boldsymbol{H}_1 \boldsymbol{R}_1 \boldsymbol{H}_1^*} \boldsymbol{R}_1 \boldsymbol{H}_1^* \boldsymbol{H}_1 \boldsymbol{R}_1^*, \tag{37}$$

where $\boldsymbol{R}_1$ is the covariance of $\boldsymbol{\theta}_1$, and is normally assumed to be $\boldsymbol{R}_{\boldsymbol{\theta}}$ as defined in (17). In practice, the disturbed $\bar{\boldsymbol{P}}_1$ is derived from the incorrect noise variance $\bar{r}_v$ and the incorrect initial state variance $\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}$. In most cases, one cannot ensure that $\bar{\boldsymbol{P}}_1$ satisfies the conditions in [94] and [95], i.e, $\bar{\boldsymbol{P}}_1 \lesseqgtr \boldsymbol{P}_1$ or $\bar{\boldsymbol{P}}_1 = \alpha \boldsymbol{P}_1$. Although it is known from [93] that the proposed synchronizer is stable as long as $\bar{\boldsymbol{P}}_1$ is bounded, the question still remains as to whether or not the actual error covariance $\breve{\boldsymbol{P}}_k$ converges to the real error covariance $\boldsymbol{P}_k$ for any $\bar{\boldsymbol{P}}_1 > 0$. The following theorem establishes the convergence of the proposed synchronizer with incorrect channel parameters.

Theorem 2: Consider the proposed synchronizer operating on time-invariant channels with incorrect statistical modeling parameters $\bar{r}_v$, $\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{P}}_1$. If we assume that each symbol is selected uniformly from the $\mathcal{M}$-ary alphabet, the error variances of the time and phase estimates converges *in probability* to the error variances of the time and phase estimates obtained with correct statistical modeling parameters.

Proof: For time-invariant channels with $\boldsymbol{w}_k = 0$, estimating $\boldsymbol{\epsilon}_k$ with the Kalman filter in (24) is mathematically equivalent to estimating $\boldsymbol{\theta}$ with the linear MMSE estimator in (22) at the k-th time interval. Therefore, the error covariances of $\hat{\boldsymbol{\epsilon}}_k$ and $\hat{\boldsymbol{\theta}}_k$ are identical. This is also true in case that the incorrect noise variance $\bar{r}_v$ and the incorrect state covariance $\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}$ are used. Because of the difficulties in analyzing the convergence of the quadratic Riccati equation in (35), we hereafter turn our attention to the linear MMSE estimator.

With statistical modeling errors, the linear MMSE estimator is

$$\bar{\hat{\boldsymbol{\theta}}}_k = \bar{\boldsymbol{R}}_{\boldsymbol{\theta}} \boldsymbol{S}_k^* \left(\boldsymbol{S}_k \bar{\boldsymbol{R}}_{\boldsymbol{\theta}} \boldsymbol{S}_k^* + \bar{r}_v \boldsymbol{I}_k \right)^{-1} \boldsymbol{z}_k, \tag{38}$$

and the actual error covariance $\breve{\boldsymbol{M}}_k$ is

$$\begin{aligned}
\breve{\boldsymbol{M}}_k &= \mathbb{E}\{|\boldsymbol{\theta}-\bar{\hat{\boldsymbol{\theta}}}_k|^2\},\\
&= \mathbb{E}\{|\boldsymbol{\theta}-\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*\left(\boldsymbol{S}_k\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*+\bar{r}_v\boldsymbol{I}_k\right)^{-1}(\boldsymbol{S}_k\boldsymbol{\theta}+\boldsymbol{v}_k)|^2\},\\
&= \boldsymbol{R}_{\boldsymbol{\theta}}-\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*\left(\boldsymbol{S}_k\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*+\bar{r}_v\boldsymbol{I}_k\right)^{-1}\boldsymbol{S}_k\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}-\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*\left(\boldsymbol{S}_k\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*+\bar{r}_v\boldsymbol{I}_k\right)^{-1}\boldsymbol{S}_k\boldsymbol{R}_{\boldsymbol{\theta}}\\
&\quad+\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*\left(\boldsymbol{S}_k\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*+\bar{r}_v\boldsymbol{I}_k\right)^{-1}(\boldsymbol{S}_k\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*+r_v\boldsymbol{I}_k)\left(\boldsymbol{S}_k\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*+\bar{r}_v\boldsymbol{I}_k\right)^{-1}\boldsymbol{S}_k\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}, \quad (39)
\end{aligned}$$

under the assumption that $\boldsymbol{\theta}$ is independent of $\boldsymbol{v}_k$. With $\boldsymbol{E}_k=\boldsymbol{S}_k\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*\left(\boldsymbol{S}_k\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*+\bar{r}_v\boldsymbol{I}_k\right)^{-1}$, the nonnegative definite symmetric matrix $\boldsymbol{S}_k\breve{\boldsymbol{M}}_k\boldsymbol{S}_k^*$ reduces to

$$\begin{aligned}
\boldsymbol{S}_k\breve{\boldsymbol{M}}_k\boldsymbol{S}_k^* &= \boldsymbol{S}_k\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*-\boldsymbol{S}_k\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*\boldsymbol{E}_k^*-\boldsymbol{E}_k\boldsymbol{S}_k\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*+\boldsymbol{E}_k\left(\boldsymbol{S}_k\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*+r_v\boldsymbol{I}_k\right)\boldsymbol{E}_k^*,\\
&= \left(\boldsymbol{I}_k-\boldsymbol{E}_{\boldsymbol{s}}\right)\boldsymbol{S}_k\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*\left(\boldsymbol{I}_k-\boldsymbol{E}_{\boldsymbol{s}}^*\right)+r_v\boldsymbol{E}_{\boldsymbol{s}}\boldsymbol{E}_{\boldsymbol{s}}^*. \quad (40)
\end{aligned}$$

Given $\boldsymbol{R}_{\boldsymbol{\theta}},\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}>0$, both $\boldsymbol{S}_k\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*$ and $\boldsymbol{S}_k\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*$ have at most N_f nonzero eigenvalues, which are denoted as $\{\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_{N_f}\}$ and $\{\bar{\lambda}_1\geqslant\bar{\lambda}_2\geqslant\cdots\geqslant\bar{\lambda}_{N_f}\}$, respectively. By using the rightmost of the following inequalities [93]

$$\lambda_{\min}(\boldsymbol{A})\mathrm{trace}(\boldsymbol{B})\leqslant\mathrm{trace}(\boldsymbol{AB})\leqslant\lambda_{\max}(\boldsymbol{A})\mathrm{trace}(\boldsymbol{B}),\quad\forall\ \boldsymbol{B}\geqslant 0, \quad (41)$$

with $\lambda_{\min}(\boldsymbol{A})$ and $\lambda_{\max}(\boldsymbol{A})$ being the minimum and maximum eigenvalues of $\boldsymbol{A}$, one can show

$$\begin{aligned}
\mathrm{trace}(\boldsymbol{S}_k\breve{\boldsymbol{M}}_k\boldsymbol{S}_k^*) &= \mathrm{trace}\left(\boldsymbol{S}_k\boldsymbol{R}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*\left(\boldsymbol{I}_k-\boldsymbol{E}_{\boldsymbol{s}}^*\right)\left(\boldsymbol{I}_k-\boldsymbol{E}_{\boldsymbol{s}}\right)\right)+\mathrm{trace}\left(r_v\boldsymbol{E}_{\boldsymbol{s}}\boldsymbol{E}_{\boldsymbol{s}}^*\right),\\
&\leqslant \lambda_1\,\mathrm{trace}\left(\left(\boldsymbol{I}_k-\boldsymbol{E}_{\boldsymbol{s}}^*\right)\left(\boldsymbol{I}_k-\boldsymbol{E}_{\boldsymbol{s}}\right)\right)+\mathrm{trace}\left(r_v\boldsymbol{E}_{\boldsymbol{s}}\boldsymbol{E}_{\boldsymbol{s}}^*\right),\\
&= \sum_{i=1}^{N_f}\frac{\lambda_1\bar{r}_v^2}{(\bar{\lambda}_i+\bar{r}_v)^2}+\sum_{i=1}^{N_f}\frac{\bar{\lambda}_i^2 r_v}{(\bar{\lambda}_i+\bar{r}_v)^2}. \quad (42)
\end{aligned}$$

The proof of the third line in (42) is elaborated as follows. Denote the single value decomposition of $\boldsymbol{S}_k\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*$ as $\boldsymbol{S}_k\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*\boldsymbol{V}_k\bar{\boldsymbol{D}}_k\boldsymbol{V}_k^*$, where $\boldsymbol{V}_k$ is a unitary matrix and $\bar{\boldsymbol{D}}_k$ is a diagonal matrix with the diagonal entries being $\{\bar{\lambda}_1,\cdots,\bar{\lambda}_{N_f},0,0,\cdots\}$. Hence, the matrix $\boldsymbol{E}_k$ can also be expressed as

$$\begin{aligned}
\boldsymbol{E}_k &= \boldsymbol{S}_k\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*\left(\boldsymbol{S}_k\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}\boldsymbol{S}_k^*+\bar{r}_v\boldsymbol{I}_k\right)^{-1},\\
&= \boldsymbol{V}_k\bar{\boldsymbol{D}}_k\boldsymbol{V}_k^*\left(\boldsymbol{V}_k\bar{\boldsymbol{D}}_k\boldsymbol{V}_k^*+\bar{r}_v\boldsymbol{V}_k\boldsymbol{V}_k^*\right)^{-1},\\
&= \boldsymbol{V}_k\Big(\bar{\boldsymbol{D}}_k(\bar{\boldsymbol{D}}_k+\bar{r}_v\boldsymbol{I}_k)^{-1}\Big)\boldsymbol{V}_k^*. \quad (43)
\end{aligned}$$

Now the trace of $\boldsymbol{E}_k\boldsymbol{E}_k^*$ can be computed as

$$\begin{aligned}\text{trace}\,(\boldsymbol{E}_k\boldsymbol{E}_k^*) &= \text{trace}\,\Big(\boldsymbol{V}_k(\bar{\boldsymbol{D}}_k(\bar{\boldsymbol{D}}_k+\bar{r}_v\boldsymbol{I}_k)^{-1})^2\boldsymbol{V}_k^*\Big),\\ &= \text{trace}\,\Big((\bar{\boldsymbol{D}}_k(\bar{\boldsymbol{D}}_k+\bar{r}_v\boldsymbol{I}_k)^{-1})^2\Big),\\ &= \sum_{i=1}^{N_f}\frac{\bar{\lambda}_i^2}{(\bar{\lambda}_i+\bar{r}_v)^2}. \end{aligned} \tag{44}$$

Similarly, the matrix $\boldsymbol{I}_k-\boldsymbol{E}_k$ can be rewritten as

$$\begin{aligned}\boldsymbol{I}_k-\boldsymbol{E}_k &= \boldsymbol{V}_k\Big(\boldsymbol{I}_k-\bar{\boldsymbol{D}}_k(\bar{\boldsymbol{D}}_k+\bar{r}_v\boldsymbol{I}_k)^{-1}\Big)\boldsymbol{V}_k^*,\\ &= \bar{r}_v\boldsymbol{V}_k(\bar{\boldsymbol{D}}_k+\bar{r}_v\boldsymbol{I}_k)^{-1}\boldsymbol{V}_k^*, \end{aligned} \tag{45}$$

and the trace of $(\boldsymbol{I}_k-\boldsymbol{E}_k)(\boldsymbol{I}_k-\boldsymbol{E}_k)^*$ is

$$\begin{aligned}\text{trace}\,\Big((\boldsymbol{I}_k-\boldsymbol{E}_k)(\boldsymbol{I}_k-\boldsymbol{E}_k)^*\Big) &= \text{trace}\,\Big((\bar{r}_v(\bar{\boldsymbol{D}}_k+\bar{r}_v\boldsymbol{I}_k)^{-1})^2\Big),\\ &= \sum_{i=1}^{N_f}\frac{\bar{r}_v^2}{(\bar{\lambda}_i+\bar{r}_v)^2}. \end{aligned} \tag{46}$$

With (44) and (46), it is straightforward to arrive at (42).

If the sampling instant is adaptively updated at each time epoch, then $\boldsymbol{S}_k$ is full rank with probability 1 for $k\geqslant N_f$,[2] and so is $\boldsymbol{S}_k\boldsymbol{R}_{\boldsymbol{\theta}}^{1/2}$, where $\boldsymbol{R}_{\boldsymbol{\theta}}^{1/2}$ is the lower triangular square root factor of $\boldsymbol{R}_{\boldsymbol{\theta}}$, i.e., $\boldsymbol{R}_{\boldsymbol{\theta}}^{1/2}\left(\boldsymbol{R}_{\boldsymbol{\theta}}^{1/2}\right)^*=\boldsymbol{R}_{\boldsymbol{\theta}}$. Moreover, if we further assume that each of the M^L linearly independent CPM signals appears in $\boldsymbol{S}_k$ with equal probability, then $\boldsymbol{S}_k^*\boldsymbol{S}_k$ becomes a strictly diagonally dominant matrix with each diagonal increasing linearly in k, as k is sufficiently large. As a result, there exist constants $d_s, d_t\in(0,+\infty)$ such that

$$\lim_{k\to+\infty}\mathbb{P}(\{d_s\leqslant\mu_i/k\leqslant d_t\})=1,\quad\forall\ 1\leqslant i\leqslant N_f, \tag{47}$$

where $\{\mu_1\geqslant\mu_2\geqslant\cdots\geqslant\mu_{N_f}\}$ are the eigenvalues of $\left(\boldsymbol{S}_k\boldsymbol{R}_{\boldsymbol{\theta}}^{1/2}\right)^*\boldsymbol{S}_k\boldsymbol{R}_{\boldsymbol{\theta}}^{1/2}$. Note that $\mu_i=\lambda_i$ for $1\leqslant i\leqslant N_f$. Similarly, one can show that there exist constants $\bar{d}_s, \bar{d}_t\in(0,+\infty)$ such that

$$\lim_{k\to+\infty}\mathbb{P}(\{\bar{d}_s\leqslant\bar{\lambda}_i/k\leqslant\bar{d}_t\})=1,\quad\forall\ 1\leqslant i\leqslant N_f. \tag{48}$$

[2]The argument for this claim is similar to the proof of *Lemma 1*.

We are now ready to deliver our main result. For every $\varepsilon \geqslant 0$,

$$\begin{aligned}
\lim_{k\to+\infty} \mathbb{P}\left(\mathrm{trace}\Big(\boldsymbol{S}_k(\breve{\boldsymbol{M}}_k - \boldsymbol{M}_k)\boldsymbol{S}_k^*\Big) \geqslant \varepsilon\right) &\leqslant \lim_{k\to+\infty} \mathbb{P}\left(\sum_{i=1}^{N_f}\left[\frac{\lambda_1 \bar{r}_v^2 + \bar{\lambda}_i^2 r_v}{(\bar{\lambda}_i + \bar{r}_v)^2} - \frac{\lambda_i r_v}{\lambda_i + r_v}\right] \geqslant \varepsilon\right), \\
&\leqslant \lim_{k\to+\infty} \mathbb{P}\left(N_f\left[\frac{k d_t \bar{r}_v^2 + k^2 \bar{d}_t^2 r_v}{(k\bar{d}_t + \bar{r}_v)^2} - \frac{k d_s r_v}{k d_s + r_v}\right] \geqslant \varepsilon\right), \\
&= \lim_{k\to+\infty} \mathbb{P}\left(\frac{k^2 \mathcal{C}_2 + k\,\mathcal{C}_1}{(k\bar{d}_t + \bar{r}_v)^2 (k d_s + r_v)} \geqslant \varepsilon\right), \\
&= 0, \qquad (49)
\end{aligned}$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are two nonnegative constants determined by N_f, d_t, d_s, $\bar{d}_t$, r_v and $\bar{r}_v$. By employing the left inequality in (41), it can be shown that

$$\lambda_{\min}(\boldsymbol{S}_k^* \boldsymbol{S}_k)\mathrm{trace}(\breve{\boldsymbol{M}}_k - \boldsymbol{M}_k) \leqslant \mathrm{trace}\Big(\boldsymbol{S}_k(\breve{\boldsymbol{M}}_k - \boldsymbol{M}_k)\boldsymbol{S}_k^*\Big), \qquad (50)$$

which yields the more insightful result

$$\begin{aligned}
\lim_{k\to+\infty} \mathbb{P}\left(\mathrm{trace}(\breve{\boldsymbol{M}}_k - \boldsymbol{M}_k) \geqslant \varepsilon\right) &\leqslant \lim_{k\to+\infty} \mathbb{P}\left(\frac{1}{\lambda_{\min}(\boldsymbol{S}_k^* \boldsymbol{S}_k)}\mathrm{trace}\Big(\boldsymbol{S}_k(\breve{\boldsymbol{M}}_k - \boldsymbol{M}_k)\boldsymbol{S}_k^*\Big) \geqslant \varepsilon\right), \\
&= 0, \qquad (51)
\end{aligned}$$

since $\boldsymbol{S}_k^* \boldsymbol{S}_k$ is nonsingular with probability 1. Finally, it is immediate from (51) that $\forall\ \varepsilon \geqslant 0$,

$$\lim_{k\to+\infty} \mathbb{P}\left(|\mathrm{var}(\bar{\hat{\theta}}_{k,i}) - \mathrm{var}(\hat{\theta}_{k,i})| \geqslant \varepsilon\right) = 0, \quad \text{for } 1 \leqslant i \leqslant N_f. \square \qquad (52)$$

2.4 Numerical Results And Discussions

The performance of the proposed time and phase synchronization approach is evaluated by computer simulation for a few specific CPM systems on time-invariant and time-variant channels. The design issues, including the selection of frequency components in nOEEs and the multiple initializations of the employed Kalman filter, are discussed. The effect of incorrect channel modeling on the convergence of the proposed synchronizer is also investigated. Although it is possible to extend the proposed technique to its decision-directed version, data-aided synchronization is considered in the following simulations. For each system setup, both the MSE and symbol error rate (SER) are obtained by averaging over 100 runs, and a different training sequence is randomly generated in each run. Moreover, we assume that frequency offset estimation has been perfectly achieved at the receiver.

2.4.1 Synchronizer Design Issues

Recall that the essential reason that we estimate $\boldsymbol{\epsilon}_k$ rather than $\boldsymbol{\theta}_k$ within the proposed synchronizer is to avoid phase wrapping when calculating $\hat{\tau}_k$ and $\hat{\phi}_k$ from $\hat{\boldsymbol{\theta}}_k$ for large timing shifts and phase offsets. The state vector $\boldsymbol{\epsilon}_k$ carries only the synchronization errors which are usually small and, thus, it seems unnecessary to consider the effects of phase wrapping. However, the errors between estimates and true parameters are unavoidably large when the synchronizer is initialized and, thus, phase wrapping still occurs. To solve this problem, two measures are taken as follows:

1. The first exponential pulse of nOEE is devised to be the DC component, i.e., $f_1 = 0$, such that only the information of the phase offset remains in $\epsilon_{k,1}$. Consequently, the phase offset is decoupled from the time shift. Therefore, a reliable estimate of the phase offset can be always obtained whenever phase wrapping occurs.

2. The frequency of the second pulse is forced to be a small value, such that phase wrapping rarely occurs in $\angle\hat{\epsilon}_{k,2}$ as long as $\phi_k - \hat{\phi}_k$ is small. By using the timing information extracted from $\angle\hat{\epsilon}_{k,2}$, we can detect whether or not phase wrapping is present in the remaining terms of $\Delta\Phi_k$, and make a correction if it does occur. Note that f_2 cannot be too small to obtain the reliable timing estimate from $\angle\hat{\epsilon}_{k,2}$. Hence, we require that f_2 lie on $[0.1T_s, 0.3T_s]$.

With the above additional constraints on the selection of exponential pulses, the resulting signal approximation accuracy is of interest. Table 2 shows the suboptimal frequency sets obtained by the exhaustive search and the corresponding MSEs for both SnOEE and nSnOEE with three pulses. Compared to the non-constrained nOEEs, the loss of accuracy caused by the use of the constrained nOEEs is observed but is not significant. We will see later that the joint time and phase synchronization will benefit greatly from the cost we paid here. Note that even with the two additional constraints, the constrained nSnOEE is still more accurate than the non-constrained SnOEE.

Table 2: Performance of the 3-dimentional nOEEs and constrained nOEEs.

Exp. Expansions $\Longrightarrow$		SnOEE[a]	Const. SnOEE[a]	nSnOEE[b]	Const. nSnOEE[b]
4-3RC,	$\boldsymbol{f}_{opt}/T_s$	$[-0.9, 0.0, 0.9]$	$[0.0, 0.3, -0.3]$	$[0.25, 0.95, 1.0]$	$[0.0, 0.3, 1.0]$
$h=\frac{1}{2}$	NMSE[c]	9.4×10^{-2}	1.6×10^{-1}	5.9×10^{-3}	1.8×10^{-2}
GMSK,	$\boldsymbol{f}_{opt}/T_s$	$[-0.45, 0.0, 0.45]$	$[0.0, 0.3, -0.3]$	$[-0.05, 0.25, 0.55]$	$[0.0, 0.25, 0.55]$
$BT_s=0.3$	NMSE[c]	7.3×10^{-4}	1.3×10^{-3}	1.8×10^{-4}	1.9×10^{-4}

[a] For SnOEE ν_{opt} is optimized by a coarse exhaustive search with grid spacing $\triangle\nu = 0.05$.

[b] For nSnOEE $\boldsymbol{f}_{opt}$ is estimated by a coarse exhaustive search with grid spacing $\triangle f_i = 0.05$.

[c] NPulses and NMSE stand for number of pulses and normalized MSE, respectively.

Alert readers might realize that the estimate of the phase offset is still unreliable at the very beginning of the synchronization even when using our phase-time-decoupled exponential expansions. As a result, one is still unable to detect and correct the harmful phase wrapping within $\Delta\Phi$. An effective approach for overcoming this difficulty is to keep the initial timing estimate unchanged during the first N_1 symbol intervals until the estimate of the phase offset becomes reliable. Afterwards, we reinitialize the Kalman filter by setting $\bar{\boldsymbol{P}}_{N_1+1} = \boldsymbol{R}_{\boldsymbol{\theta}}$, and start to update the time shift and the phase offset simultaneously. We call this method a double-initialized Kalman filtering. Note that at the second initialization, the incorrect initial state error covariance is employed, yielding a slower convergence rate. In order to improve the convergence performance of the proposed synchronizer, we reinitialize the Kalman filter again at the N_2+1-st time epoch. Since both the timing and phase information are usually quite reliable at this time, the corresponding actual error variances are small. Therefore, it is reasonable to approximate $\boldsymbol{P}_{N_2+1}$ with a more accurate estimate $\alpha_R\boldsymbol{R}_{\boldsymbol{\theta}}$, where $0.01\leqslant\alpha_R\leqslant0.1$.

Fig. 1 compares the performance of the double- and triple-initialized synchronizers. A CPM system with 4-ary symbols, 3RC phase shaping pulse, and the modulation index $h_{\text{idx}} = 1/2$ is considered on a time-invariant AWGN channel. The initial time shift and phase offset are $0.4T_s$ and 0.6π, respectively. The double and triple initialization time epochs are $N_1=40$ and $N_2 = 80$, respectively. For clarity, only the normalized MSE of the timing estimate

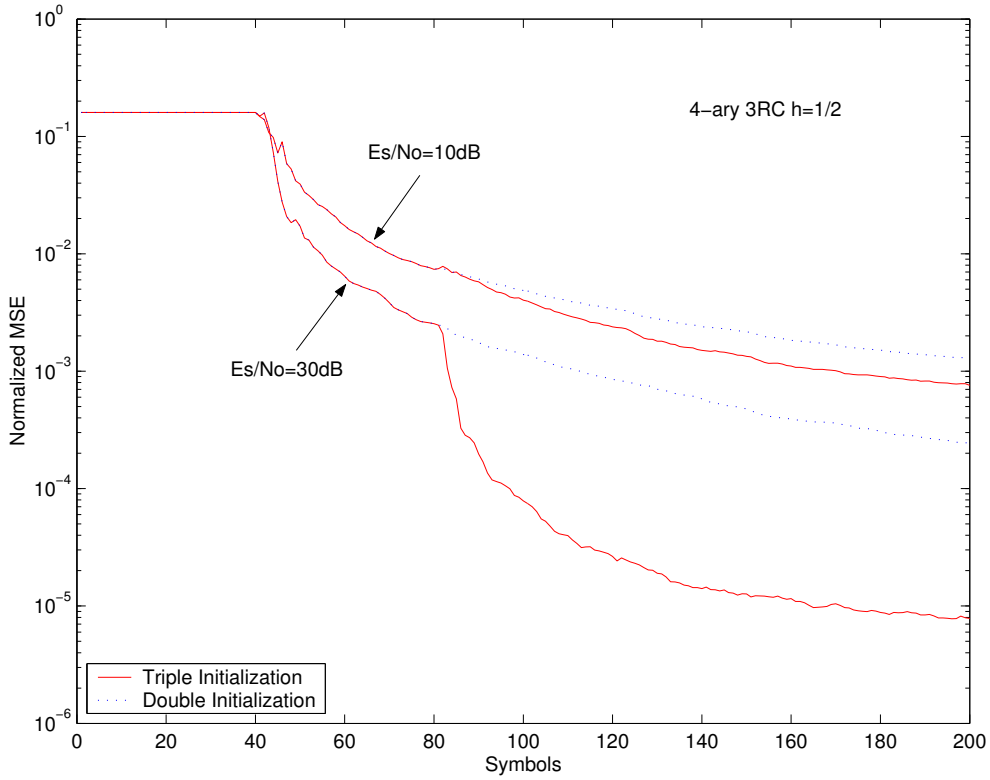


Figure 1: Normalized MSE performance of the proposed synchronizer with double- and triple-initialized Kalman filtering. The constrained nSnOEE with 3 pulses is used. The normalized time variance is defined as $\mathrm{var}(\hat{\tau}_k)/T_s^2$.

is plotted in Fig. 1. At different values of E_s/N_o, both the double- and triple-initialized synchronizers exhibit a robust performance in the presence of salient initial parameter errors, while the triple-initialized synchronizer exhibits superior convergence performance. With an increase in E_s/N_o, the performance improvement from using the triple-initialized Kalman filtering is more notable. This is expected since the effect of the incorrect initial error covariance dominates the convergence performance when the channel noise diminishes. At low or medium values of E_s/N_o, the convergence rate is mainly determined by the channel noise and, thus, the performance gain resulted from the adjustment in the initial error covariance is smaller.

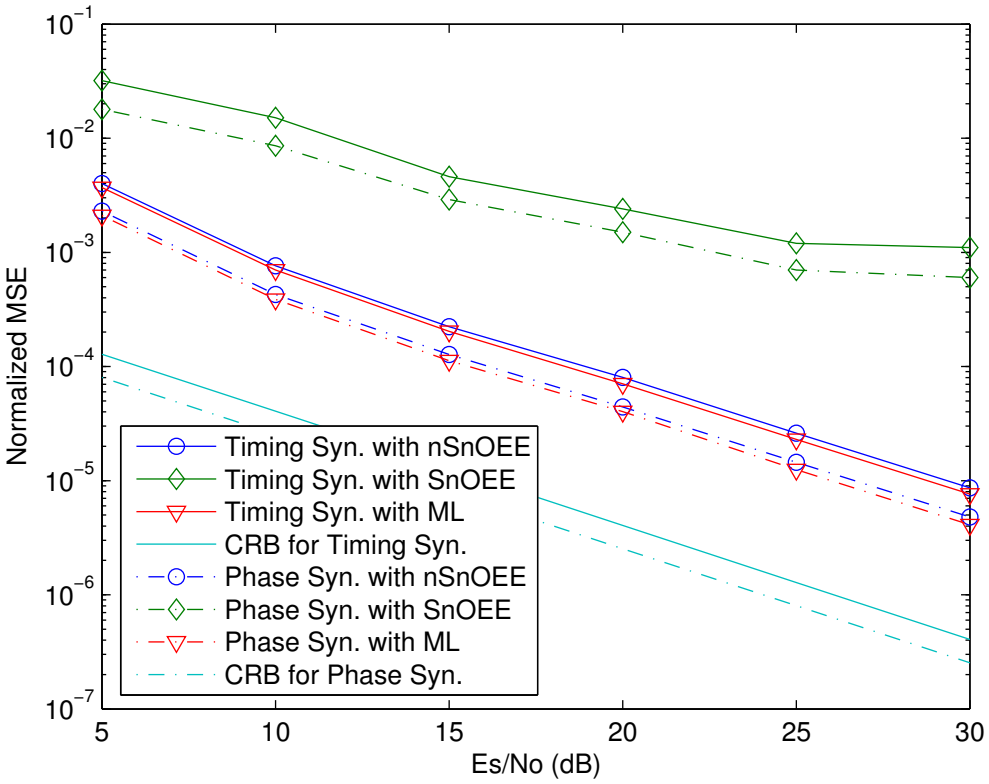


Figure 2: Normalized MSE performance of the proposed synchronizer with constrained SnOEE and nSnOEE for a 4-ary-3RC CPM system with $h_{\text{idx}} = 1/2$. The training length is $N_u = 200$. A triple-initialized Kalman filter with $N_1 = 40$ and $N_2 = 80$ is employed. The normalized time and phase variance are defined as $\text{var}(\hat{\tau}_k)/T_s^2$ and $\text{var}(\hat{\phi}_k)/\pi^2$.

2.4.2 MSE Performance of the Proposed Synchronizer

The MSE performance of the proposed triple-initialized synchronizer with the constrained 3-dimensional SnOEE and nSnOEE is shown in Fig. 2 and 3 for the 4-ary-3RC CPM system with $h_{\text{idx}} = 1/2$. The suboptimal frequency sets in Table II are used accordingly. The length of training sequence is $N_u = 200$ in Fig. 2 and $N_u = 20$ in Fig. 3. The time-invariant additive white Gaussian noise (AWGN) channel with the same setup as in Fig. 1 is considered. The performance of the data-aided ML synchronizer in [6] as well as the Cramer-Rao bounds (CRB) of joint time and phase estimation [96],[97] are also shown for convenience of comparison. For the ML synchronizer, we assume the full-complexity FEP in which *derivative* matched filters are used. Within computation of CRBs, the unknown parameters are assumed to be deterministic, which is statistically equivalent to the uniform *a-priori*

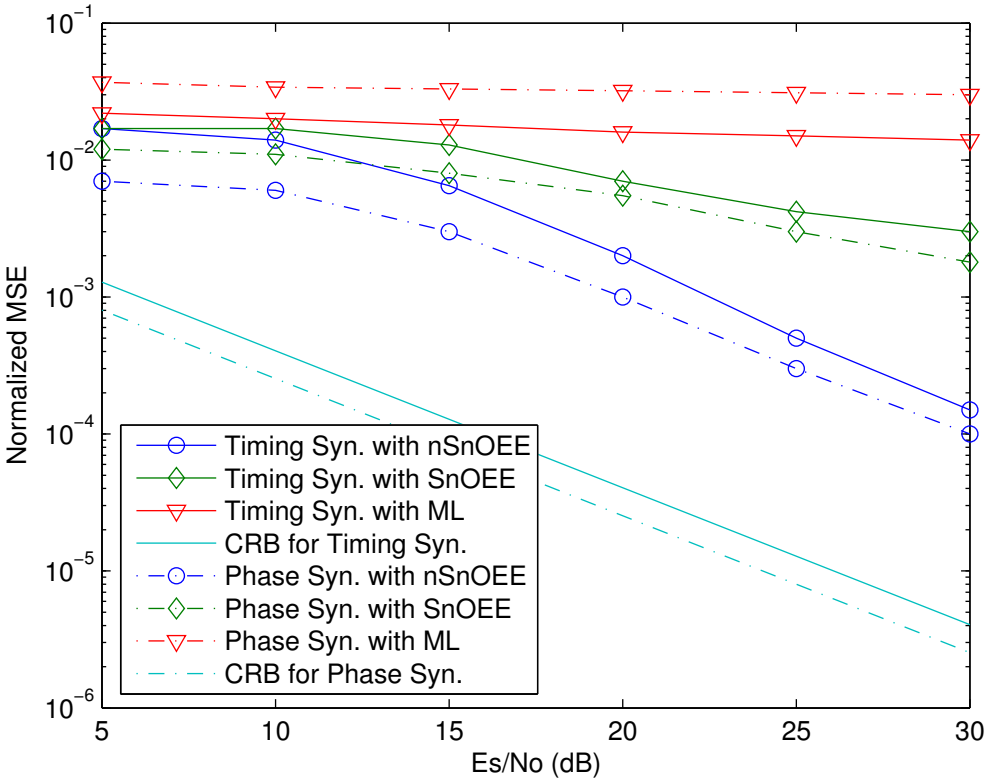


Figure 3: Normalized MSE performance of the proposed synchronizer with constrained SnOEE and nSnOEE for a 4-ary-3RC CPM system with $h_{\text{idx}} = 1/2$. The training length is $N_u = 20$. A triple-initialized Kalman filter with $N_1 = 5$ and $N_2 = 12$ is employed. The normalized time and phase variance are defined as $\text{var}(\hat{\tau}_k)/T_s^2$ and $\text{var}(\hat{\phi}_k)/\pi^2$.

assumption for timing and phase errors in the synchronizer design.

It is shown in Fig. 2 and 3 that the proposed synchronizer with the constrained nSnOEE is able to acquire the unknown time and phase shift whenever $N_u = 200$ or $N_u = 20$. The performance gap between the constrained nSnOEE and SnOEE is significant in both figures because of the coarse signal approximation of the latter, and a noticeable error floor is observed at the high E_s/N_o when SnOEE is used. Since the traditional ML synchronizer uses the full-complexity FEP, it performs slightly better than the proposed synchronizer with its much less complex FEP, when sufficient training is available, e.g, $N_u = 200$. In some applications, such as burst mode transmission, only a short preamble can be used to acquire unknown parameters. In Fig. 3 it is shown that when $N_u = 20$ the parameter acquisition fails with the traditional ML synchronizer, while it succeeds with the proposed

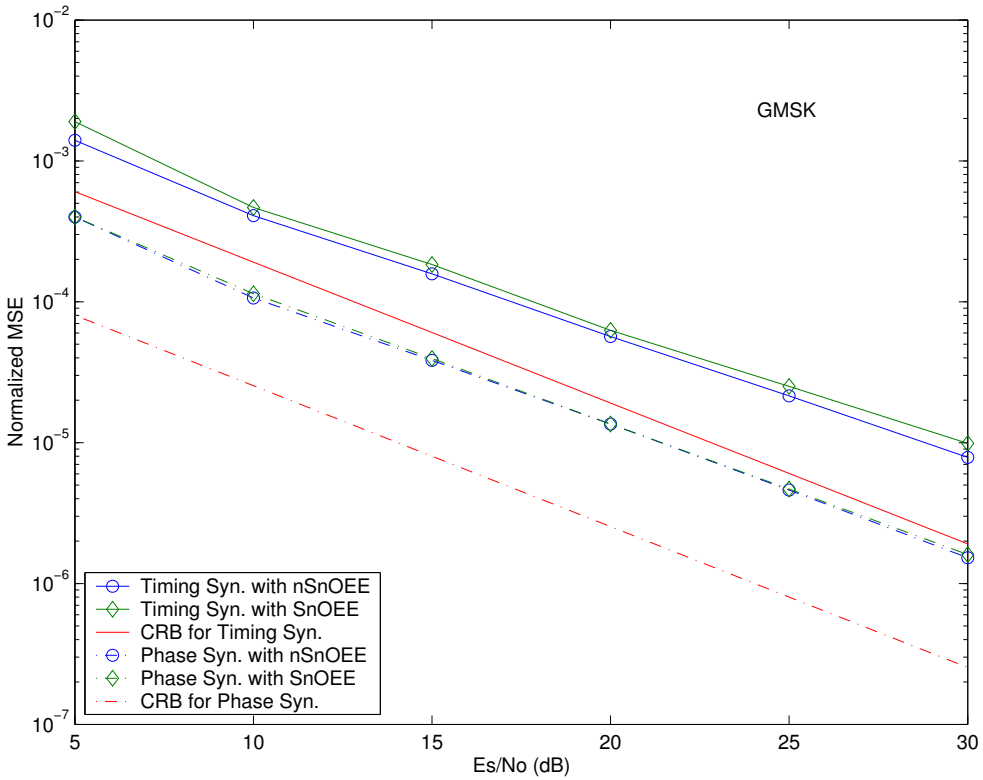


Figure 4: Normalized MSE performance of the proposed synchronizer with constrained SnOEE and nSnOEE for GMSK with $BT_s = 0.30$. The training length is $N_u = 200$. A triple-initialized Kalman filter with $N_1 = 40$ and $N_2 = 80$ is employed. The normalized time and phase variance are defined as $\mathrm{var}(\hat{\tau}_k)/T_s^2$ and $\mathrm{var}(\hat{\phi}_k)/\pi^2$.

nSnOEE-based receiver owing to the fast converging Kalman filter.

The MSE performance of the proposed synchronizer for GMSK with $BT_s = 0.3$ is shown in Fig. 4, when sufficient training is assumed, e.g., $N_u = 200$. In contrast to 4-ary-3RC CPM, the synchronizer performance with the two nOEEs differs only slightly for GMSK although their signal approximation accuracies are considerably different. An intuitive explanation for this phenomenon is that the synchronization performance is dominated by other factors, such as interpolation errors, ISI or incorrect modeling, once the signal approximation is sufficiently accurate. Our simulations have shown that the performance penalty resulted from imperfect signal approximation of the nOEEs is negligible once the normalized MSE of the approximate signals is less than 1×10^{-3}. This observation implies that it is unnecessary to invoke the nOEEs with more than three pulses to synchronize most CPM signals.

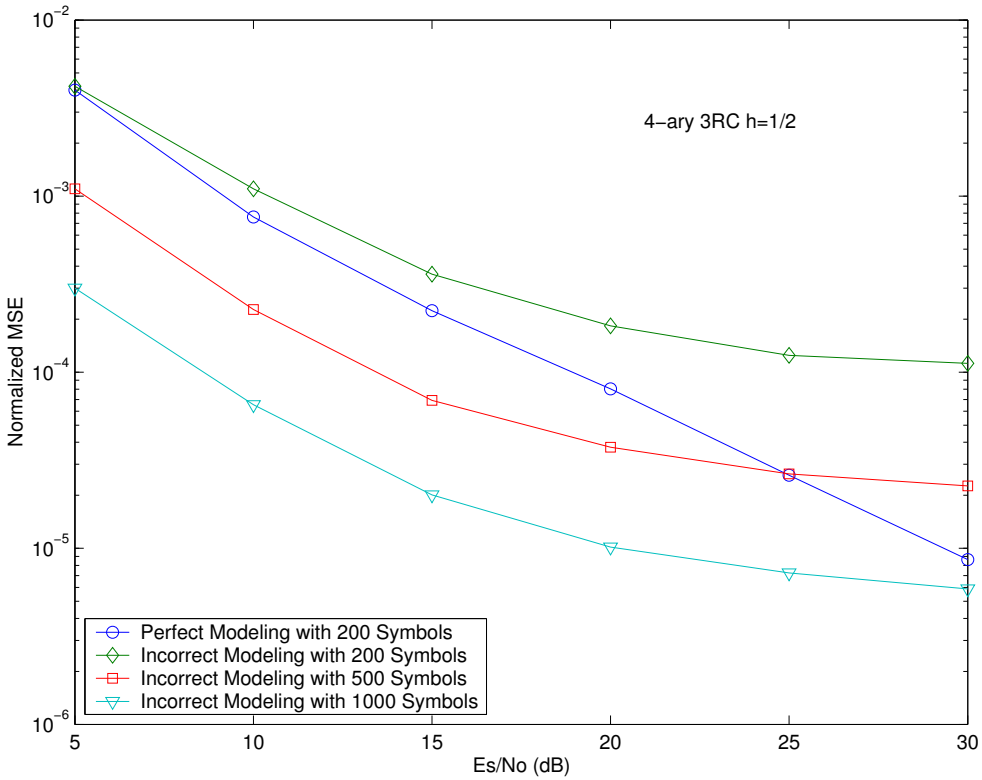


Figure 5: Normalized timing MSE performance of the proposed synchronizer with incorrect modeling errors for a 4-ary-3RC CPM system with $h_{\text{idx}} = 1/2$. A triple-initialized Kalman filter with $N_1 = 40$ and $N_2 = 80$ is employed. The normalized time variance is defined as $\text{var}(\hat{\tau}_k)/T_s^2$.

2.4.3 Effect of Modeling Errors

The effect of statistical modeling errors on the performance of the proposed synchronizer is investigated via simulations. Again the 4-ary-3RC CPM system with $h_{\text{idx}} = 1/2$ is considered through this subsection. Fig. 5 shows the normalized time error variances of the proposed synchronizer with incorrect channel modeling and with different data lengths. The following scenario is assumed in the figure: 1) the unknown time and phase shift are time-invariant; 2) the probability distributions of the desired parameters are unknown and, thus, the covariance of the state vector $\boldsymbol{\theta}$ is simply set to $\bar{\boldsymbol{R}}_{\boldsymbol{\theta}} = \boldsymbol{I}_{N_f}$; 3) the variance of measurement noise r_v is obtained from one shot estimation as $\bar{r}_v = 0.3162\, E_s$, i.e., $\overline{E_s/N_o} = 5\,\text{dB}$, and then used for subsequent processing regardless of whether or not the actual noise variance changes with time. When the actual E_s/N_o is 5 dB, the performance penalty resulted from modeling errors is essentially due to the incorrect state covariance $\bar{\boldsymbol{R}}_{\boldsymbol{\theta}}$, and is negligible in

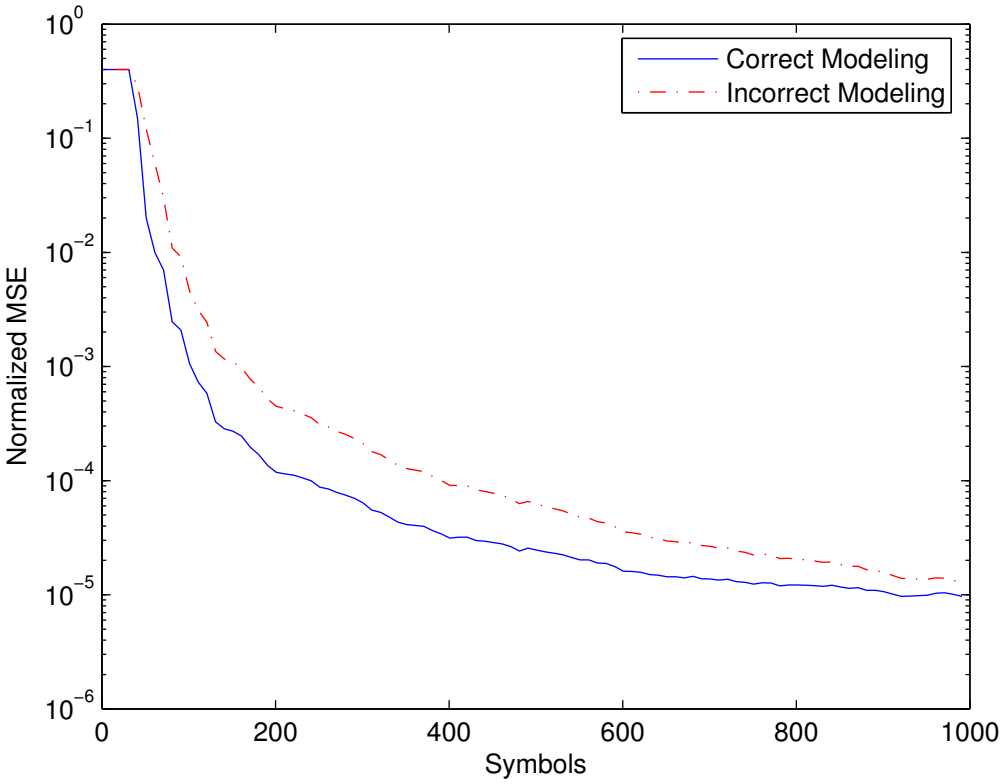


Figure 6: Convergence performance of the proposed synchronizer with incorrect modeling errors for a 4-ary-3RC CPM system with $h_{\text{idx}} = 1/2$ on a time-invariant channel with $E_s/N_o = 20\,\text{dB}$. A triple-initialized Kalman filter with $N_1 = 40$ and $N_2 = 80$ is employed. The normalized time variance is defined as $\text{var}(\hat{\tau}_k)/T_s^2$.

Fig. 5. Along with the increase of E_s/N_o, the ratio of the assumed noise variance to the actual noise variance rises correspondingly, and the resulting performance loss gets worse until it flattens at 30 dB. However, the error floor can be significantly reduced by increasing the training duration. Hence, on time-invariant channels the modeling errors do not yield irreducible errors but only slow down the synchronizer convergence, which confirms the conclusion of *Theorem* 2.

Fig. 6 suggests a different perspective to understand *Theorem* 2. As in Fig. 5, we still assume $\bar{\boldsymbol{R}}_{\boldsymbol{\theta}} = \boldsymbol{I}_{N_f}$ and $\overline{E_s/N_o} = 5\,\text{dB}$, while the true E_s/N_o is 20 dB. Although the MSE performance of the synchronizer with incorrect modeling is worse than that of the synchronizer with correct modeling, both of them converge gradually and the performance gap between them diminishes with an increase of training length.

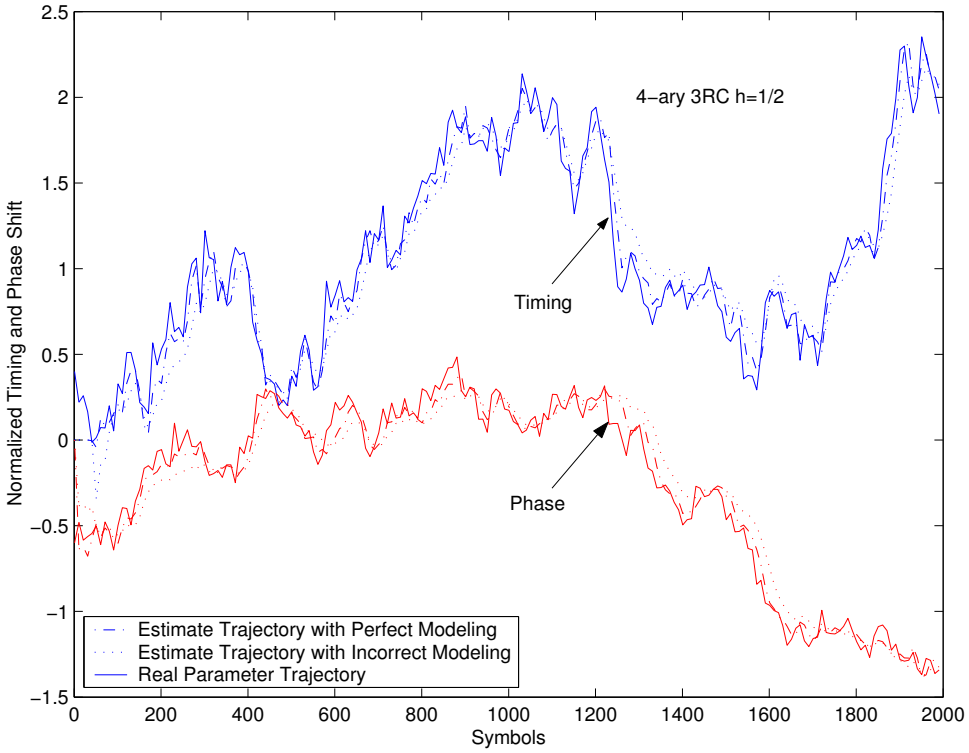


Figure 7: Estimate trajectories of the proposed synchronizer with correct and incorrect modeling errors with $E_s/N_o = 15\,\text{dB}$. Triple-initialization with $N_1 = 40$ and $N_2 = 80$.

Fig. 7 demonstrates the robustness of the proposed synchronizer with or without modeling errors on time-variant channels. The unknown time and phase jitters are generated independently by two random walks with root-mean-squares of $0.05T_s$ and 5^o, respectively. When the channel modeling is perfect, Fig. 7 shows that the proposed synchronizer is robust in acquiring and tracking the rapidly time-varying parameters. Even when the modeling errors appear as $\bar{\boldsymbol{R}}_{\boldsymbol{\theta}} = \boldsymbol{I}_{N_f}$, $\bar{r}_v = 10r_v$, and $\bar{\boldsymbol{Q}} = 10\boldsymbol{Q}$, the proposed synchronizer is still robust although a slight performance degradation is observed.

2.4.4 BER Performance

The performance of the proposed FEP and synchronizer is evaluated in Fig. 8 in terms of BER. An AWGN channel with the time-invariant parameters $\tau = 0.4T_s$ and $\phi = 0.6\pi$ is considered. When both timing and phase shifts are ideally known at the receiver, the performance loss caused by the use of nSnOEE-based suboptimal FEP is negligible compared to the receiver with full-complexity FEP. When the synchronization parameters are

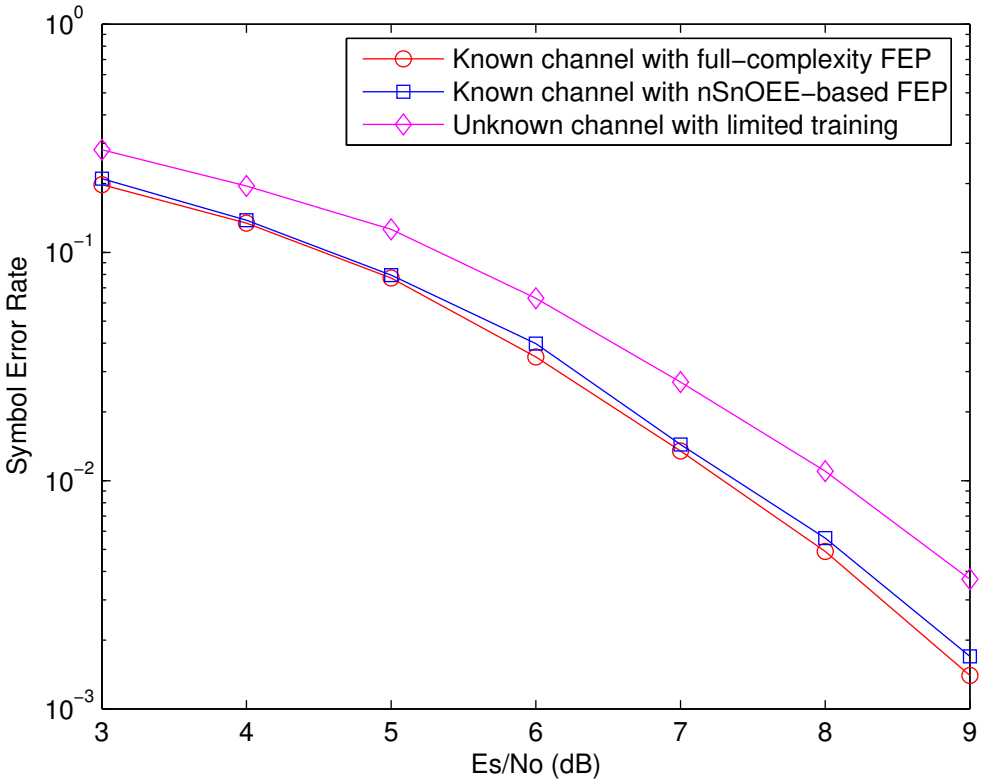


Figure 8: BER performance of the proposed receiver for a 4-ary-3RC CPM system with $h_{\text{idx}} = 1/2$. The nSnOEE-based FEP is used. A triple-initialized Kalman filter with $N_1 = 40$ and $N_2 = 80$ is employed.

unknown at the receiver, a preamble is used to acquire them, and the estimates obtained at the end of acquisition are then used through the following demodulation as constants. We assume in the simulation that the length of the preamble is 100 and the length of an information sequence is 900. Hence the resulting power overhead due to parameter training is 0.46 dB. It is shown in Fig. 8 that over time-invariant channels the total performance loss due to the proposed data-aided synchronization is approximately 0.8 dB, 0.3 dB of which is caused by imperfect parameter estimation.

Decision-directed synchronization is more power efficient than data-aided synchronization and is particularly suitable when unknown parameters are time-variant. However, it was reported in [18] that the decision-directed synchronizer cannot synchronize partial-response CPM signals regardless of the format of applied time and phase estimators, because

of the existence of false locks. The extension of the proposed synchronizer to its decision-directed fashion is non-trivial and is tackled in Chapter V, where we introduce a mixed method for CPM turbo synchronization that can effectively prevent the occurrence of false locks.

CHAPTER III

ITERATIVE PHASE SYNCHRONIZATION FOR SERIALLY CONCATENATED CPM

The serial concatenation of a convolutional code and a CPM modulator yields a type of serially concatenated convolutional codes and, thus, near Shannon-limit performance can be achieved by performing iterative demodulation/decoding [7],[8],[33]. However, these systems typically operate at a low SNR and, therefore, reliable synchronization is challenging, especially for fast time-variant fading channels. In this chapter, we develop a novel adaptive soft-input soft-output (A-SISO) module for MAP symbol detection with parameter uncertainty. Based on this scheme, a family of *fixed-interval* A-SISO (FI-A-SISO) algorithms are utilized to implement blind iterative phase synchronization for SCCPM. For SCCPM systems, the *forward-only* FI-A-SISO algorithms are shown to be much more robust than the *forward-backward* FI-A-SISO algorithms in acquiring and tracking a time-varying carrier phase. The reason has to do with rotational invariance of CPM signals. This result can be extended to any rotationally invariant convolutional coded system with an unknown carrier phase. The main results in this chapter can be found in [98] .

3.1 Innovations-Based MAP Estimation

3.1.1 A Near-Optimal Evaluation for Innovations-Based MAP Estimation

We assume that a length-N_u symbol sequence $\mathbf{u}_1^{N_u} = [u_1, u_2, \cdots, u_{N_u}]$ is transmitted and must be detected at a receiver. The criterion for MAP symbol detection is

$$\hat{u}_k = \arg \max_{\tilde{u}_k \in \mathcal{U}} \ \mathbb{P}\left(\tilde{u}_k | \mathbf{y}_1^{N_u}\right), \tag{53}$$

where $\tilde{u}_k$ and $\hat{u}_k$ represent the hypothesis and estimate of u_k, respectively, $\mathbf{y}_1^{N_u} = [\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_{N_u}]$ is a set of discrete observables at output of the front-end processor and consists of sufficient

statistics for optimal detection[2], $\mathbb{P}(\tilde{u}_k|\mathbf{y}_1^{N_u})$ is the *a-posteriori* probability (APP) of the tentative information symbol at the k-th time epoch[3], and $\mathcal{U}$ is an M-ary symbol alphabet. Applying Bayes' rule and the principle of total probability, and omitting a normalizing constant, the criterion in (53) can be rewritten as

$$\hat{u}_k = \arg\max_{\tilde{u}_k \in \mathcal{U}} \sum_{(\tilde{\mathbf{u}}_1^{N_u}:\tilde{u}_k)} p(\mathbf{y}_1^{N_u}|\tilde{\mathbf{u}}_1^{N_u})\mathbb{P}(\tilde{\mathbf{u}}_1^{N_u}), \tag{54}$$

where $(\tilde{\mathbf{u}}_1^{N_u} : \tilde{u}_k)$ indicates that the summation is over all possible symbol sequences with the hypothesized symbol $\tilde{u}_k$ at the k-th interval.

If unknown and random parameters exist, the conditional probability density function $p(\mathbf{y}_1^{N_u}|\tilde{\mathbf{u}}_1^{N_u})$ in (54) can be computed by averaging as follows,

$$p(\mathbf{y}_1^{N_u}|\tilde{\mathbf{u}}_1^{N_u}) = \int_{\mathbf{\Theta}} p(\mathbf{y}_1^{N_u}|\tilde{\mathbf{u}}_1^{N_u}, \tilde{\boldsymbol{\theta}}_1^{N_u}) p(\tilde{\boldsymbol{\theta}}_1^{N_u}) d\tilde{\boldsymbol{\theta}}_1^{N_u}, \tag{55}$$

where $\mathbf{\Theta}$ is a parameter space, $\tilde{\boldsymbol{\theta}}_1^{N_u}$ is the hypothesis of $\boldsymbol{\theta}_1^{N_u} = [\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, \cdots, \boldsymbol{\theta}_{N_u}^T]^T$ that represents an unknown time-variant parameter sequence independent of the information symbols, $\boldsymbol{\theta}_i$ is a column vector with dimension L_θ and L_θ is the number of unknown parameters at each time interval. The evaluation of (55) requires knowledge of the *a priori* statistics of $\boldsymbol{\theta}_1^{N_u}$, which is unlikely to be available in practice. Further simplification can be taken by assuming $\boldsymbol{\theta}_1^{N_u}$ is uniformly distributed. In this case, a reasonable approximation of $p(\mathbf{y}_1^{N_u}|\tilde{\mathbf{u}}_1^{N_u})$ is

$$p(\mathbf{y}_1^{N_u}|\tilde{\mathbf{u}}_1^{N_u}) \approx p\left(\mathbf{y}_1^{N_u}|\tilde{\mathbf{u}}_1^{N_u}, \hat{\boldsymbol{\theta}}_1^{N_u}(\mathbf{y}_1^{N_u}, \tilde{\mathbf{u}}_1^{N_u})\right), \tag{56}$$

where $\hat{\boldsymbol{\theta}}_1^{N_u}(\mathbf{y}_1^{N_u}, \tilde{\mathbf{u}}_1^{N_u})$ is an estimate of $\boldsymbol{\theta}_1^{N_u}$ given the observation record $\mathbf{y}_1^{N_u}$ and each sequence hypothesis $\tilde{\mathbf{u}}_1^{N_u}$. The combination of (54) and (56) suggests a two-step approximate solution to the original problem in (53), i.e, first estimating unknown parameters for each possible data sequence and then evaluating the APP of information symbols. In this thesis, we shall focus on the linear least-squares estimates (LLSEs) of $\boldsymbol{\theta}_1^{N_u}$. With a

[2]Note that multiple samples could be extracted in each symbol interval, i.e, $\mathbf{y}_k = [y_{k,1}, y_{k,2}, \cdots, y_{k,J}]$. When unknown parameters exist, a proper frond-end processor is required to generate sufficient statistics for optimal joint estimation and detection [101].

[3]In the rest of the thesis, $\mathbb{P}(\cdot)$ is used to represent the probability mass function of a discrete random variable/vector while $p(\cdot)$ representing the probability density function of a continuous random variable/vector.

known discrete-time state-space linear model, $\hat{\boldsymbol{\theta}}_1^{N_u}(\mathbf{y}_1^{N_u}, \tilde{\mathbf{u}}_1^{N_u})$ can be computed by using a class of innovations-based estimators [90]. In the following, we simplify the notation of $\hat{\boldsymbol{\theta}}_1^{N_u}(\mathbf{y}_1^{N_u}, \tilde{\mathbf{u}}_1^{N_u})$ with $\hat{\boldsymbol{\theta}}_1^{N_u}$.

Let $\hat{\boldsymbol{\theta}}_{i|N_u}$ denote the LLSE of $\boldsymbol{\theta}_i$ given the entire observation record $\mathbf{y}_1^{N_u}$. Moreover, let $\hat{\boldsymbol{\theta}}_{i|i}$ and $\hat{\boldsymbol{\theta}}_{i|i-1}$ be the LLSEs of $\boldsymbol{\theta}_i$ given $\mathbf{y}_1^i$ and $\mathbf{y}_1^{i-1}$, respectively. The computation of $\hat{\boldsymbol{\theta}}_{i|N_u}$, $\hat{\boldsymbol{\theta}}_{i|i}$ and $\hat{\boldsymbol{\theta}}_{i|i-1}$ is called smoothing, forward filtering and forward prediction, respectively. Similarly, we can define the backward filtered and predicted LLSEs of $\boldsymbol{\theta}_i$ as $\hat{\boldsymbol{\theta}}_{i|i}^b$ and $\hat{\boldsymbol{\theta}}_{i|i+1}^b$ given the observation $\mathbf{y}_i^{N_u}$ and $\mathbf{y}_{i+1}^{N_u}$, respectively. As shown in Appendix A, the smoothed estimate can be represented as

$$\hat{\boldsymbol{\theta}}_{i|N_u} = \hat{\boldsymbol{\theta}}_{i|i} - \boldsymbol{C}_{i|i}(\boldsymbol{C}_{i|i} + \boldsymbol{C}_{i|i+1}^b)^{-1}(\hat{\boldsymbol{\theta}}_{i|i} - \hat{\boldsymbol{\theta}}_{i|i+1}^b), \tag{57}$$

where $\boldsymbol{C}_{i|i}$ and $\boldsymbol{C}_{i|i+1}^b$ are the covariance matrices of $\hat{\boldsymbol{\theta}}_{i|i}-\boldsymbol{\theta}_i$ and $\hat{\boldsymbol{\theta}}_{i|i+1}^b-\boldsymbol{\theta}_i$, respectively. Note that the estimate $\hat{\boldsymbol{\theta}}_{i|i}$ (or $\hat{\boldsymbol{\theta}}_{i|i+1}^b$) can be computed by using a forward (or backward) Kalman recursion. The estimation of $\hat{\boldsymbol{\theta}}_{i|i}$ (or $\hat{\boldsymbol{\theta}}_{i|i+1}^b$) is based on information extracted from the past and present (or future) received signals. Furthermore, the *smoothed* estimate $\hat{\boldsymbol{\theta}}_{i|N_u}$ can be viewed as a linear combination of the forward filtered and backward predicted estimates, such that all past, present and future information are included in the current parameter update. As a result, the smoothed estimator always outperforms the forward (or backward) predicted and filtered estimators in the sense of minimizing MSE.

Observe from (54) that the entire observation record is used to evaluate the APP of a tentative information symbol at the k-th interval. Letting $\hat{\boldsymbol{\theta}}_1^{N_u} = \{\hat{\boldsymbol{\theta}}_{i|N_u}\}_{i=1}^{N_u}$ suggests a near-optimal solution to joint estimation and detection problems, since all information is considered in both estimation and detection. Correspondingly, the MAP criterion becomes

$$\hat{u}_k = \arg\max_{\tilde{u}_k \in \mathcal{U}} \sum_{(\tilde{\mathbf{u}}_1^{N_u}:\tilde{u}_k)} p\left(\mathbf{y}_1^{N_u} | \tilde{\mathbf{u}}_1^{N_u}, \{\hat{\boldsymbol{\theta}}_{i|N_u}\}_{i=1}^{N_u}\right) \mathbb{P}\left(\tilde{\mathbf{u}}_1^{N_u}\right). \tag{58}$$

The use of the LLSEs of $\boldsymbol{\theta}_1^{N_u}$ other than $\hat{\boldsymbol{\theta}}_1^{N_u}\{\hat{\boldsymbol{\theta}}_{i|N_u}\}_{i=10}^{N_u}$ leads to a family of innovations-based MAP estimation approaches.

3.1.2 Suboptimal Algorithms for Innovations-Based MAP Estimation

Assume that the considered signaling scheme can be represented by a Markov chain. Define $\mathcal{S}$ as a set of all possible transition states, and s_{k+1} as the ending state of a sequence $\mathbf{u}_1^k$. Let t_k be the transition branch connecting states s_k and s_{k+1}. Then, with some simple manipulations, we can rewrite (58) in the split form as follows,

$$\hat{u}_k = \arg\max_{\tilde{u}_k \in \mathcal{U}} \sum_{(t_k:\tilde{u}_k)} \Big\{ p\left(\mathbf{y}_k | t_k, \hat{\boldsymbol{\theta}}_{k|N_u}\right) \mathbb{P}\left(t_k\right) \sum_{(\tilde{\mathbf{u}}_1^{k-1}:s_k)} p\left(\mathbf{y}_1^{k-1} | \tilde{\mathbf{u}}_1^{k-1}, \{\hat{\boldsymbol{\theta}}_{i|N_u}\}_{i=1}^{k-1}\right) \mathbb{P}\left(\tilde{\mathbf{u}}_1^{k-1}\right) \sum_{(\tilde{\mathbf{u}}_{k+1}^{N_u}:s_{k+1})} p\left(\mathbf{y}_{k+1}^{N_u} | \tilde{\mathbf{u}}_{k+1}^{N_u}, \{\hat{\boldsymbol{\theta}}_{i|N_u}\}_{i=k+1}^{N_u}\right) \mathbb{P}\left(\tilde{\mathbf{u}}_{k+1}^{N_u}\right) \Big\}. \quad (59)$$

Replacing the smoothed LLSEs $\{\hat{\boldsymbol{\theta}}_{i|N_u}\}_{i=1}^{N_u}$ in (59) with the predicted LLSEs $\{\hat{\boldsymbol{\theta}}_{i|i-1}\}_{i=1}^{k}$ and $\{\hat{\boldsymbol{\theta}}_{i|i+1}^{b}\}_{i=k+1}^{N_u}$ yields the original *forward-backward* MAP estimation in [40]. However, the linear predictor does not fully exploit available information in the entire observation record. To remedy this, a correction term $b(\cdot)$, called a binding factor, is further introduced to partially include the dependence of the past, present and future estimates in the current symbol detection. The resulting MAP criterion is

$$\hat{u}_k = \arg\max_{\tilde{u}_k \in \mathcal{U}} \sum_{(t_k:\tilde{u}_k)} \Big\{ p\left(\mathbf{y}_k | t_k, \hat{\boldsymbol{\theta}}_{k|k-1}\right) \mathbb{P}\left(t_k\right) b(\cdot) \sum_{(\tilde{\mathbf{u}}_1^{k-1}:s_k)} p\left(\mathbf{y}_1^{k-1} | \tilde{\mathbf{u}}_1^{k-1}, \{\hat{\boldsymbol{\theta}}_{i|i-1}\}_{i=1}^{k-1}\right) \mathbb{P}\left(\tilde{\mathbf{u}}_1^{k-1}\right) \sum_{(\tilde{\mathbf{u}}_{k+1}^{N_u}:s_{k+1})} p\left(\mathbf{y}_{k+1}^{N_u} | \tilde{\mathbf{u}}_{k+1}^{N_u}, \{\hat{\boldsymbol{\theta}}_{i|i+1}^{b}\}_{i=k+1}^{N_u}\right) \mathbb{P}\left(\tilde{\mathbf{u}}_{k+1}^{N_u}\right) \Big\}, \quad (60)$$

where $b(\cdot)$ is determined by the difference between the forward and backward estimates of $\boldsymbol{\theta}_k$, and it vanishes when the two estimates are identical. The close-form expression of $b(\cdot)$ can be found in [40] and [99]. When unknown parameters are fixed through the entire observation, we can show in Appendix B that the MAP criteria in (59) and (60) are equivalent although the different parameter estimates are used in the calculation of APP. When unknown parameters are time-variant, the two criteria are no longer equivalent and may lead to different decisions.

Another MAP estimation method, proposed in [100], can be viewed as an intermediate

step relating the criteria in (59) and (60), which is given by[4]

$$\hat{u}_k = \arg\max_{\tilde{u}_k \in \mathcal{U}} \sum_{\mathcal{A}} \Big\{ p\big(\mathbf{y}_{k-Q}^{k+Q} | \tilde{\mathbf{u}}_{k-Q}^{k+Q}, \{\hat{\boldsymbol{\theta}}_{i|N_u}\}_{i=k-Q}^{k+Q}\big) \mathbb{P}\left(\tilde{\mathbf{u}}_{k-Q}^{k+Q}\right) \sum_{(\tilde{\mathbf{u}}_1^{k-Q-1}: s_{k-Q})} p\left(\mathbf{y}_1^{k-Q-1} | \tilde{\mathbf{u}}_1^{k-Q-1}, \{\hat{\boldsymbol{\theta}}_{i|i-1}\}_{i=1}^{k-Q-1}\big) \mathbb{P}\left(\tilde{\mathbf{u}}_1^{k-Q-1}\right) \sum_{(\tilde{\mathbf{u}}_{k+Q+1}^{N_u}: s_{k+Q})} p\big(\mathbf{y}_{k+Q+1}^{N_u} | \tilde{\mathbf{u}}_{k+Q+1}^{N_u}, \{\hat{\boldsymbol{\theta}}_{i|i+1}^{b}\}_{k+Q+1}^{i=N_u}\big) \mathbb{P}(\tilde{\mathbf{u}}_{k+Q+1}^{N_u}) \Big\}, \tag{61}$$

where the smoothed estimates are used to compute the APP of a truncated portion of each hypothesized information sequence while the predicted estimates are used for the remaining APP calculation, Q is called a *completion parameter* indicating the truncation length, $\mathcal{A} = (\tilde{\mathbf{u}}_{k-Q}^{k+Q} : s_{k-Q}, s_{k+Q}, \tilde{u}_k)$ represents all possible data sequences starting from the state s_{k-Q} and ending at the state s_{k+Q} and with the trial symbol $\tilde{u}_k$. With an increment of Q, the criterion in (61) gradually approaches the criterion in (59). Also, letting $Q = 0$, the criterion in (61) is approximately equal to the criterion in (60).

The exact evaluation of the above MAP criteria in (59)$\sim$(61) requires an exhaustive search among all possible symbol sequences. Given the size of the symbol alphabet, M, the computation complexity is of order $\mathcal{O}(M^{N_u})$, which makes the detection impractical even for a small alphabet size and a short data packet length. Hence reduced-complexity suboptimal algorithms are needed.

Following similar tree pruning rules in the PSP approach [37], the two inner summations in the MAP split-form criteria can be updated recursively as in the BCJR algorithm [38], i.e.,

$$\alpha(s_{k+1}) = \sum_{(\tilde{\mathbf{u}}_1^k : s_{k+1})} p\left(\mathbf{y}_1^k | \tilde{\mathbf{u}}_1^k, \hat{\boldsymbol{\theta}}_1^k\right) \mathbb{P}\left(\tilde{\mathbf{u}}_1^k\right) \approx \sum_{s_k \in \mathcal{S}} \alpha(s_k) \gamma(s_k, s_{k+1}), \tag{62a}$$

$$\beta(s_{k+1}) = \sum_{(\tilde{\mathbf{u}}_{k+1}^{N_u} : s_{k+1})} p\left(\mathbf{y}_{k+1}^{N_u} | \tilde{\mathbf{u}}_{k+1}^{N_u}, \hat{\boldsymbol{\theta}}_{k+1}^{N_u}\right) \mathbb{P}\left(\tilde{\mathbf{u}}_{k+1}^{N_u}\right) \approx \sum_{s_{k+2} \in \mathcal{S}} \beta(s_{k+2}) \gamma(s_{k+1}, s_{k+2}), \tag{62b}$$

where $\alpha(s_k)$ and $\beta(s_k)$ are the forward and backward state metrics, respectively, and $\gamma(s_k, s_{k+1}) = p\left(\mathbf{y}_k | s_k, s_{k+1}, \hat{\boldsymbol{\theta}}_k\right) \mathbb{P}\left(s_{k+1} | s_k\right)$ is the branch metric. The approximation in (62) results from survivor selection within parameter estimation recursions. The detailed

[4]The description of the MAP criterion in (61) is slightly different from its original representation in [100], where a sliding observation window is applied on both estimation and detection.

proof of (62) is presented in Appendix C. Given the number of transition states, $|\mathcal{S}|$, the complexity of this PSP-based implementation is of order $\mathcal{O}(M|\mathcal{S}|N_u)$.

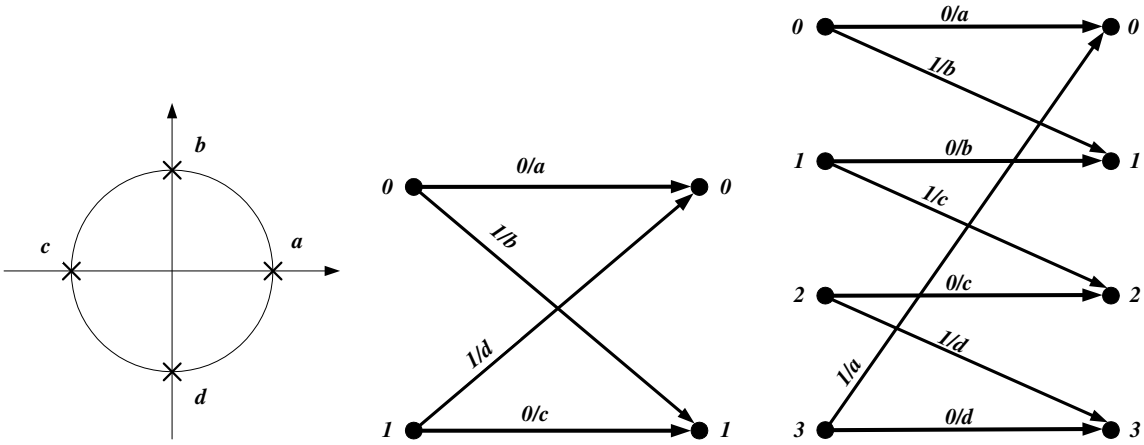
We denote the suboptimal A-SISO algorithms derived from the MAP criteria in (59)$\sim$(61) as A-SISO-Smooth, A-SISO-Bind, and A-SISO-Truncate, respectively. Recall that within A-SISO-Bind and A-SISO-Truncate the predicted forward and backward estimates are used in the forward and backward state metric recursions accordingly, while the smoothed estimates are used for both forward and backward metric recursions in A-SISO-Smooth. It is noteworthy that in A-SISO-Smooth the refined statistics owing to linear smoothing are propagated through the trellis, resulting in improved performance in both estimation and detection. In the case of time-invariant unknown parameters, although the MAP criteria in (59) and (60) are equivalent, using the smoothed estimates can lead to more reliable survivor selection and, thus, the superiority of A-SISO-Smooth over A-SISO-Bind.

3.2 CPM and RIC Codes

3.2.1 Design of RIC Codes

RIC coding is a realization of trellis-coded modulation. Given a signal constellation, we may construct a trellis having rotational invariance, such that symbol detection can tolerate certain phase errors. The design rules for RIC codes presented [51], [52], [57] and [59], are summarized as follows.

For each transition $s_k \rightarrow s_{k+1}$, we denote the associated output signal set as U, in which each signal element can be constructed from information symbols. We also assume there exists a one-to-one function $f_{\mathcal{S}} : \mathcal{S} \rightarrow \mathcal{S}$, such that the transition $f_{\mathcal{S}}(s_k) \rightarrow f_{\mathcal{S}}(s_{k+1})$ is valid in the given trellis. Let V be the signal set corresponding to the transition $f_{\mathcal{S}}(s_k) \rightarrow f_{\mathcal{S}}(s_{k+1})$. If V can also be obtained by rotating U by an angle ψ, we say that the trellis (or code) is rotationally invariant to a phase shift of ψ. Let the signal element $\mathcal{X}_U$ in U be the rotated version of the signal element $\mathcal{X}_V$ in V. If the two sets of input information symbols which correspond to the matched signal element pair $\{\mathcal{X}_U, \mathcal{X}_V\}$ happen to be same, the symbols can be correctly detected even when the phase error ψ exists. If a trellis is transparent to phase shifts $\psi_1, \cdots, \psi_l$ simultaneously, the corresponding state rotation functions $f_{\mathcal{S}}^{\psi_1}, \cdots, f_{\mathcal{S}}^{\psi_l}$



(a) Signal constellation (b) A 2-state QPSK trellis (c) A 4-state QPSK trellis

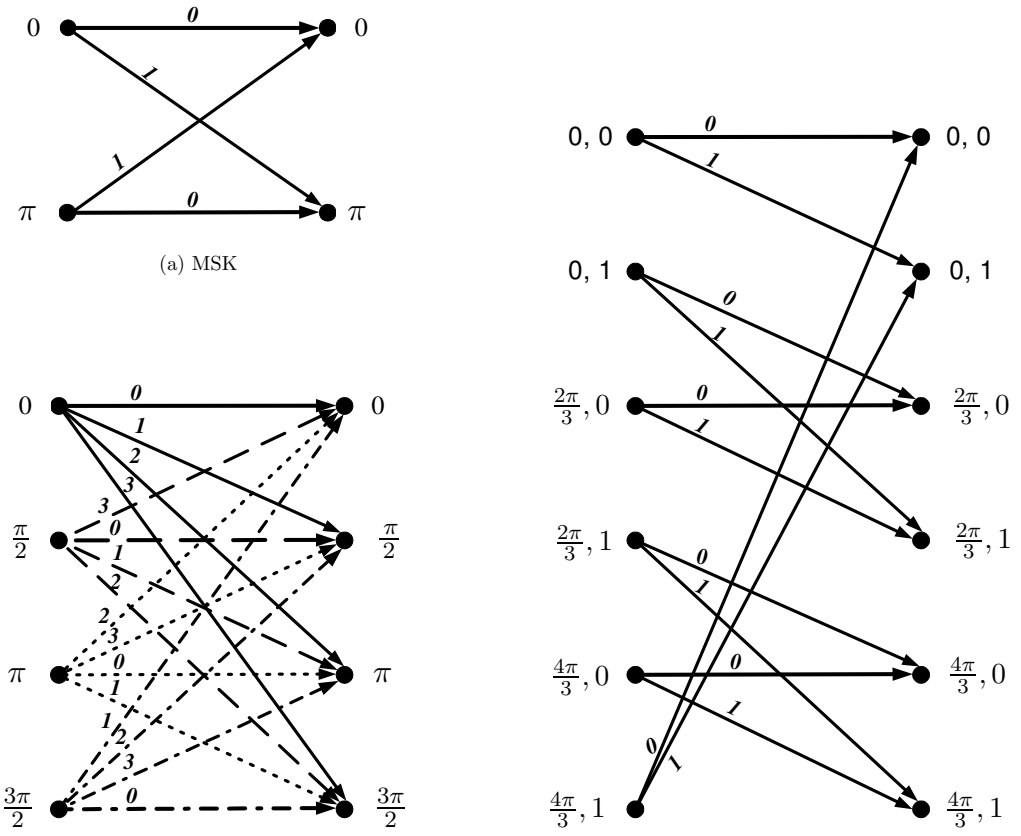
Figure 9: Trellis diagram of RIC codes.

must be determined.

For example, given the signal constellation in Fig. 9 (a), a 2-state QPSK trellis with 180° rotational invariance is shown in Fig. 9 (b). We assume that information symbols are binary and the code rate is 1/2. Denote x mod y as mod(x, y). In accordance with the state rotation function $f_{\mathcal{S}}^{180^\circ}(s_k)$=mod$(s_k+1, 2)$, the transition (0→0) under a 180° phase shift, for instance, leads to the transition (1→1); the corresponding signal pair is $\{a, c\}$ which are signal points 180° apart while the corresponding information symbols are identical. Given the same signal constellation, a more complicated 4-state QPSK trellis with 90°, 180° and 270° rotational invariance is shown in Fig. 9 (c), where three state rotation functions are defined as $f_{\mathcal{S}}^{90^\circ}(s_k)$=mod$(s_k+3, 4)$, $f_{\mathcal{S}}^{180^\circ}(s_k)$=mod$(s_k+2, 4)$ and $f_{\mathcal{S}}^{270^\circ}(s_k)$=mod$(s_k+1, 4)$, respectively. For instance, the transition $(0 \rightarrow 0)$ under a 90° phase rotation yields the transition $(3 \rightarrow 3)$; the corresponding signal pair is $\{a, d\}$ which are signal points 90° apart while the corresponding information symbols are identical.

3.2.2 Rotational Invariance of CPM Trellis

It is known that CPM exhibits rotational invariance about certain phase angles. However, the relation between RIC codes and CPM is not well studied. Recall that a CPM modulator can be represented by a recursive convolutional encoder, where the L–1-st recent symbol is



(a) MSK

(b) 4-CPFSK

(c) CPM with $h_{\text{idx}}=1/3, M=2, L=2$

Figure 10: Rotational invariance of CPM trellis.

always feedback to a modulo-2π adder to yield the next phase state. There is an intuitive relationship between the recursive structure of CPM and the required feedback structure in RIC codes. By using three examples, we show that a CPM trellis with the modulation index $h_{\text{idx}}=q/p$ can be viewed as a RIC code transparent to $p-1$ phase rotations including $\{2\pi/p, 4\pi/p, \cdots, 2\pi(p-1)/p\}$. This property is particularly valuable for phase synchronization, since the performance degradation owing to false lock, cycle slip and hangup can be easily avoided or reduced.

The first example is the MSK system. The trellis in Fig. 10 (a) corresponds to a 2-state RIC code with 180° phase rotational invariance. The state rotation function is $f_{\mathcal{S}}^{180^\circ}(s_k) = \text{mod}(s_k+\pi, 2\pi)$. For example, the transition $(0 \rightarrow 0)$ under a 180° phase rotation yields the transition $(\pi \rightarrow \pi)$. The corresponding signal pair, i.e., $x(t, \varphi_k = 0, u_k = 0)$ and $x(t, \varphi_k = \pi, u_k = 0)$, are 180° from each other. As shown in Fig. 10 (b), the 4-ary continuous

phase frequency-shift keying (CPFSK) system can be viewed as a 4-state RIC code with 90°, 180° and 270° phase rotational invariance. The state rotation functions are defined as $f_{\mathcal{S}}^{90^\circ}(s_k) = \mathrm{mod}(s_k + 3\pi/2, 2\pi)$, $f_{\mathcal{S}}^{180^\circ}(s_k) = \mathrm{mod}(s_k + \pi, 2\pi)$ and $f_{\mathcal{S}}^{270^\circ}(s_k) = \mathrm{mod}(s_k + \pi/2, 2\pi)$, respectively. In the last example shown in Fig. 10 (c), a CPM system with a rectangular pulse is considered, where $h_{\mathrm{idx}} = 1/3$, $M = 2$ and $L = 2$. Its trellis has 120° and 240° phase rotational invariance. In this case, the state s_k is denoted as $s_k = [\varphi_k, u_{k-1}]$. The state rotation functions are $f_{\mathcal{S}}^{120^\circ}([\varphi_k, u_{k-1}]) = [\mathrm{mod}(\varphi_k + 4\pi/3, 2\pi), \mathrm{mod}(u_{k-1} + 1, 2)]$ and $f_{\mathcal{S}}^{240^\circ}(s_k) = [\mathrm{mod}(\varphi_k + 4\pi/3, 2\pi), \mathrm{mod}(u_{k-1} + 1, 2)]$.

3.3 Adaptive Iterative Phase Synchronization for SCCPM Systems

3.3.1 System Description

A SCCPM system is considered as shown in Fig. 11. At the transmitter, an uncoded bit sequence is convolutionally encoded, then bit interleaved, and finally modulated by a CPM modulator. At the receiver, a proper front-end processor is required to generate sufficient statistics for MAP detection. The A-SISO CPM demodulator acts not only as an inner decoder but also as a phase synchronizer. Typical of SCCPM iterative receivers [33], [**?**], the soft information is exchanged iteratively between the A-SISO demodulator and a channel decoder. We expect that both phase estimates and information bit decisions can be further refined by this procedure.

We assume the modulated signal is propagated over a flat fading channel. The received complex low-pass envelope is

$$y(t) = h(t)x(t, \mathbf{u}_1^{N_u}) + n(t), \qquad (k-1)T_s \leqslant t \leqslant kT_s, \quad k=1, 2, \cdots, N_u, \quad (63)$$

where $h(t) = h_I(t) + jh_Q(t)$ is the time-varying complex fading gain. We also assume the amplitude of $h(t)$ is a Rayleigh-distributed process with average power normalized to unity, and the carrier phase is uniformly distributed over $[0, 2\pi]$. When Doppler effects cannot be neglected, a 2-D isotropic scattering environment is assumed where the channel gain is

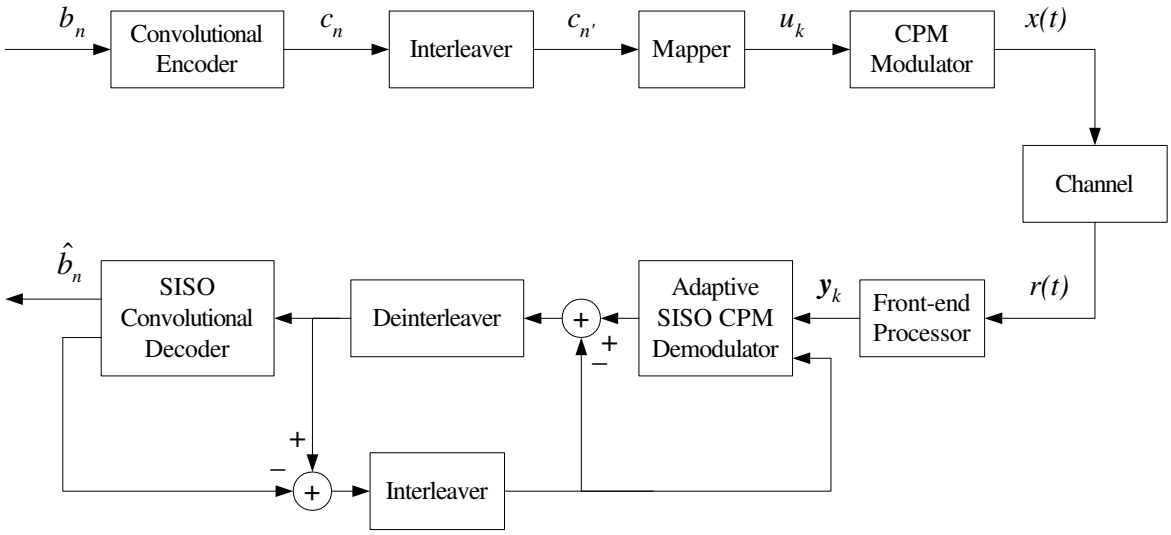


Figure 11: Transmitter and receiver structure for SCCPM systems.

correlated in time and its *normalized* autocorrelation and cross-correlation functions are

$$R_{h_I}(\tau) = R_{h_Q}(\tau) = \frac{1}{2} J_0(2\pi\tau f_m), \quad \text{and } R_{h_I,h_Q}(\tau) = 0, \tag{64}$$

in which f_m is the maximum Doppler frequency shift, and $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind. If an AWGN channel is considered, the complex channel gain reduces to $e^{j\phi(t)}$, where the carrier phase offset $\phi(t)$ usually can be represented by a phase step or a phase ramp. When random phase jitter exists, we can model $\phi(t)$ as a standard Wiener process.

When estimating unknown time-variant fading at the receiver, we simply assume that the channel gain is fixed over each symbol duration [101]. The resulting front-end processing is thus suboptimal in generating sufficient statistics for MAP detection. Assume we use the tilted-phase representation in the reminder of this chapter for simplicity of notations. Therefore, the total number of hypothesized signal waveforms within one symbol interval is pM^L, yielding a front-end processor of pM^L matched filters (MFs). Note that if two hypothesized signals only differ in the phase state, the outputs of the MFs are exactly same except for a phase offset, which is the difference between the phase states. As a result, the required number of MFs finally reduces to M^L.

3.3.2 Adaptive Iterative Phase Synchronization

Based on the innovations-based MAP estimation method, we employ a set of FI-A-SISO algorithms to implement joint phase synchronization and CPM demodulation. Compared to the SW-A-SISO algorithms with the truncation length D in [42], the FI-A-SISO algorithms are $2D$ times less computationally complex, and can achieve lower BER. However, the SW-A-SISO algorithms require less memory than our proposed FI-A-SISO algorithms for the storage of the survived estimates, and the state and branch metrics.

To perform coherent CPM demodulation, the discretized carrier phase $\phi_k = \phi((k - 1/2)T_s)$ must be acquired and tracked. The first-order forward digital phase-locked loop (DPLL), equivalent to the least-mean-squares (LMS) algorithm, is given by [6]

$$\hat{\phi}_{k|k} = \hat{\phi}_{k-1|k-1} + \eta \,\mathrm{Im}\left(\int_{(k-1)T_s}^{kT_s} y(t)x^*(t, s_k, \tilde{u}_k)e^{-j\hat{\phi}_{k-1|k-1}}dt\right), \tag{65}$$

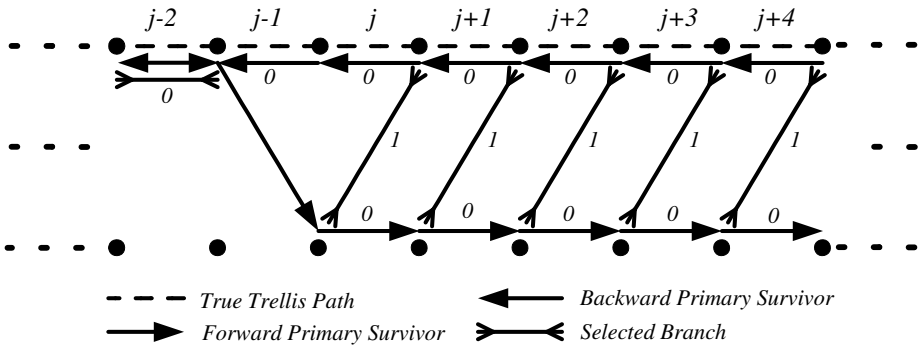
where η is the closed-loop gain (stepsize) and $\mathrm{Im}(\cdot)$ represents the imaginary part of a complex variable. Since the carrier phase is the only unknown parameter, the vector $\boldsymbol{\theta}_k$ reduces to the scalar ϕ_k. Similarly, we can obtain the backward DPLL. Moreover, the *smoothed* estimate is

$$\hat{\phi}_{k|N_u} = \frac{k}{N_u}\hat{\phi}_{k|k} + \frac{N_u - k}{N_u}\hat{\phi}^b_{k|k+1}, \tag{66}$$

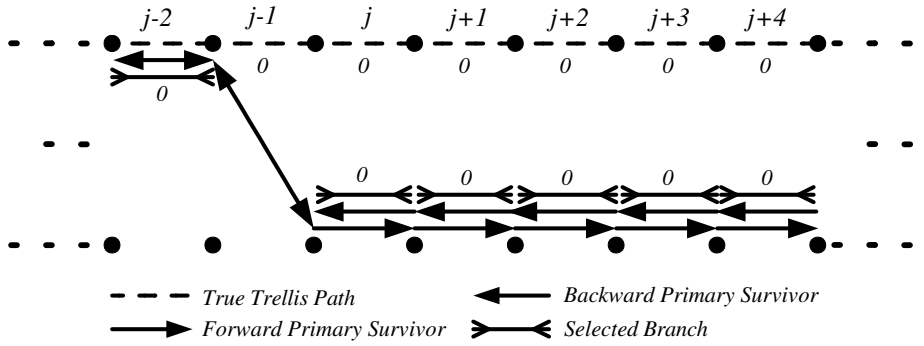
as shown in Appendix A. In the following, the notation of the *fixed-interval*, i.e, FI, is dropped for simplicity. Besides A-SISO-Smooth and A-SISO-Bind, the *forward-only* A-SISO algorithm (A-SISO-FO), is also implemented, in which the forward estimates $\{\hat{\phi}_{i|i}\}_{i=0}^{N_u}$ are used for both forward and backward state metric recursions.

3.3.3 Inherent Limitations of Bidirectional A-SISO Algorithms

Within the bidirectional A-SISO algorithms, including A-SISO-Bind and A-SISO-Smooth, etc., survivor selection is independently performed in the two directions. For non-RIC codes, the primary survivor in each direction, i.e., the path with maximum APP, always manages to merge to the real path. Therefore, the forward and backward primary survivors are almost same except at few time stages. However, this property does not hold for RIC codes in



(a) Survivors for the *forward-backward* FI-A-SISO algorithms.



(b) Survivors for the *forward-only* FI-A-SISO algorithms.

Figure 12: Survivor paths within MSK trellis.

case of joint phase estimation and detection. The possible mismatches between the forward and backward primary survivors can significantly degrade the detection performance of the bidirectional A-SISO algorithms. We start our investigation with MSK.

Recall that the MSK trellis has 180° rotational invariance. Thus, a 180° phase error does not cause a detection error. For example, let us consider the trellis in Fig. 12, where we assume that an all-zeroes symbol sequence is transmitted and no additive noise is present. Suppose that owing to some unexpected reason a 180^o phase estimation error occurs at the $j-1$-st epoch within the forward parameter estimation. Thereafter, the forward primary survivor diverges from the true path. However, the information symbols which are associated with the mistakenly selected path are still correct owing to the rotational invariance of the given trellis. We emphasize that if no further phase estimation error occurs, the selected path will never merge to the real path.

Within MSK SISO demodulation, the log-APP-ratio of the k-th information symbol is

$$\begin{aligned}\Gamma(u_k) &= \ln \frac{\mathbb{P}(u_k=0|\mathbf{y}_1^{N_u})}{\mathbb{P}(u_k=1|\mathbf{y}_1^{N_u})} \\ &= \ln\Big[\alpha(s_k=0)\gamma(s_k=0,s_{k+1}=0)\beta(s_{k+1}=0)+\alpha(s_k=\pi)\gamma(s_k=\pi,s_{k+1}=\pi)\beta(s_{k+1}=\pi)\Big] - \\ &\quad \ln\Big[\alpha(s_k=0)\gamma(s_k=0,s_{k+1}=\pi)\beta(s_{k+1}=\pi)+\alpha(s_k=\pi)\gamma(s_k=\pi,s_{k+1}=0)\beta(s_{k+1}=0)\Big] \quad (67)\end{aligned}$$

If $\Gamma(u_k)>0$, then $u_k=0$; otherwise, $u_k=1$. Recall that the forward state metric $\alpha(s_k=s)$ is the sum of the APPs of all possible symbol sequences ending with the state s at the k-th symbol interval; and a similar argument is true for the backward state metric $\beta(s_{k+1}=s)$. In Fig. 12 (a), the forward primary survivor ends with the state π for $j\leqslant k\leqslant N_u$ and, thus, the APP of the forward primary survivor is a dominant term in $\alpha(s_k=\pi)$. Also, it is highly possible that $\alpha(s_k=\pi)$ is much greater than $\alpha(s_k=0)$ when $j\leqslant k\leqslant N_u$. If we assume that no phase estimation errors occur in the backward estimation, then the backward primary survivor is the same as the true path. Consequently, $\beta(s_k=0)$ is greater than $\beta(s_k=\pi)$ with a high probability, for $1\leqslant k\leqslant N_u$. Based on the above discussions, it is highly possible that the term $\alpha(s_k=\pi)\gamma(s_k=\pi,s_{k+1}=0)\beta(s_{k+1}=0)$ dominates the evaluation of $\Lambda(u_k)$ and yields a wrong decision, i.e, $\hat{u}_k=1$, for $j\leqslant k\leqslant N_u$. In Fig. 12 (b), we assume that the forward phase estimates will be used in the backward state metric recursion and, thus, the forward and backward primary survivors are same. In this case, for $j\leqslant k\leqslant N_u$, the term $\alpha(s_k=\pi)\gamma(s_k=\pi,s_{k+1}=\pi)\beta(s_{k+1}=\pi)$ will dominate the evaluation of $\Gamma(u_k)$ with a high probability, and lead to correct decisions.

Generally speaking, for the A-SISO algorithms, the APPs of the forward and backward primary survivors provide a major contribution within the evaluation of the log-APP-ratios of information symbols. Therefore, if a branch connecting with both forward and backward primary survivors corresponds to an incorrect information symbol, then the final decision at this time stage is wrong with a high probability. For RIC codes, once primary survivors diverge from a true path, they will never come back if no further phase rotations occur. The presence of the rotational invariance in a trellis results in different forward and backward primary survivors in the *forward-backward* A-SISO algorithms, which makes them more vulnerable to phase errors than the *forward-only* A-SISO algorithms. For non-RIC codes,

primary survivors eventually merge into a true path even when a few phase rotations exist. As a result, the forward and backward primary survivors are almost same everywhere and, therefore, the *forward-backward* and *forward-only* A-SISO algorithms possess approximately same robustness to phase errors.

3.4 Numerical Results And Discussions

The performance of the innovations-based FI-A-SISO algorithms have been evaluated by extensive simulations. The serially concatenated MSK (SCMSK) system is cosnsidered. The adaptive iterative receivers are compared in terms of BER, MSE of phase estimates, and robustness to time-variant fading and phase jitter. In order to increase power efficiency, no pilots are transmitted. In the DPLL, the closed-loop gain is optimized such that BER is always minimized. At the transmitter, a pseudo-random interleaver is used, and a rate-1/2 convolutional code with generators $[5_8, 7_8]$ is employed. Therefore, the bit energy E_b is defined as $E_b = 2E_s/\log_2 M$. Moreover, we assume for simplicity that frequency synchronization, timing recovery and noise variance estimation have been perfectly achieved at the receiver.

Because of constant modulus of CPM, the effect of inaccurate fading amplitude in turbo decoding can be equivalently transformed to the famous SNR mismatch problem. It is shown in [102] that performance loss owing to the SNR mismatch is negligible when the E_b/N_o offset falls in $[-3\,\text{dB}, 6\,\text{dB}]$. For this reason we do not perform any estimation and compensation of the fading amplitude at the receiver. This is equivalent to saying that the estimate of fading amplitude is always unity although the actual fading amplitude is a Rayleigh-distributed with unit average power. Our simulation results confirm that the performance loss owing to this simplification is negligible on fading channels.

3.4.1 Performance of the FI-A-SISO Algorithms

The performance of the FI-A-SISO Algorithms is evaluated in this subsection. Note that A-SISO-FO and A-SISO-Bind can be viewed as the equivalent counterparts of the forward-only and binding FI-A-SISO algorithms in [41]; however, in [41] non-RIC codes are considered,

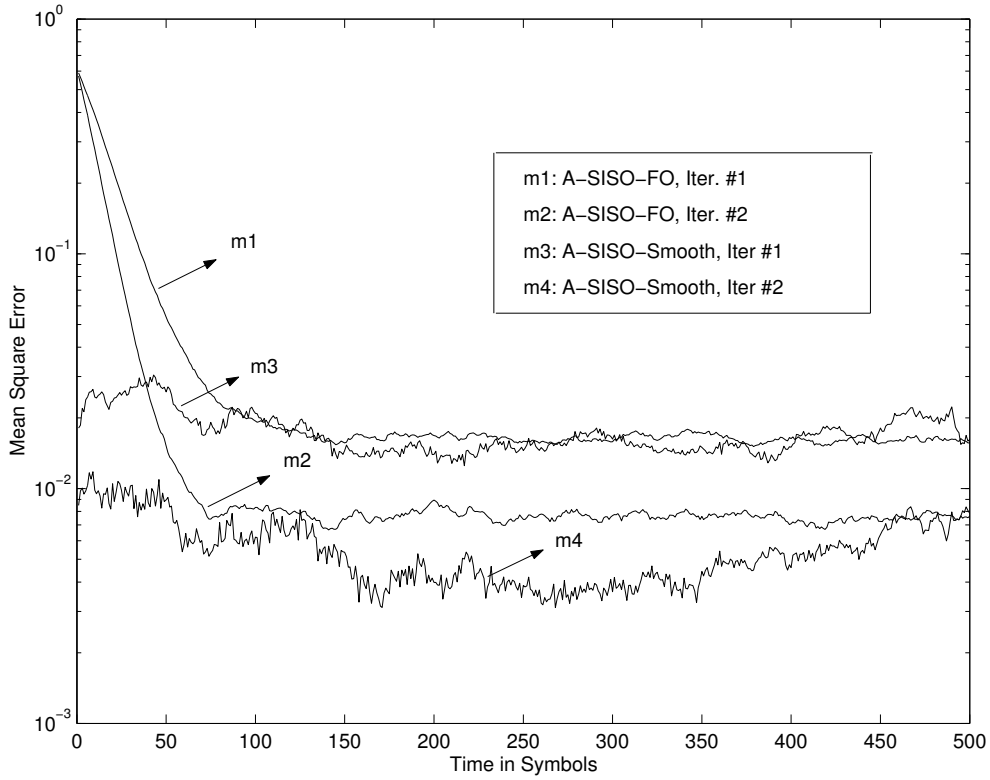


Figure 13: MSE of phase estimates for the FI-A-SISO algorithms. An SCMSK system operating on an AWGN channel with a phase step is considered. The iterative receiver stops after 2 iterations. The interleaver size is $N_u = 500$.

and pilot symbols are used to track a slow time-varying carrier phase. For A-SISO-Smooth, there is no equivalent counterpart in [40] and [41].

The MSE performance of the FI-A-SISO phase synchronizers is evaluated in Fig. 13. An AWGN channel with an unknown phase step of $\pi/4$ is assumed, and E_b/N_o is set to 5 dB such that the carrier phase is accurately tracked and no phase rotation appears during the tracking stage. Fig. 13 shows that the MSE of the phase estimates can be reduced through iterations, although the performance improvement diminishes after 2 iterations. Since more information is considered in the smoothed estimator, a smaller MSE is obtained by using A-SISO-Smooth as compared to A-SISO-FO. Although not shown in Fig. 13, it is noteworthy that A-SISO-Bind has the same MSE performance as A-SISO-FO, because the binding factor only represents an adjustment on the APP of tentative transmitted symbols, but not on the phase estimates directly.

Further BER results of the FI-A-SISO algorithms are given in Fig. 14. The performance

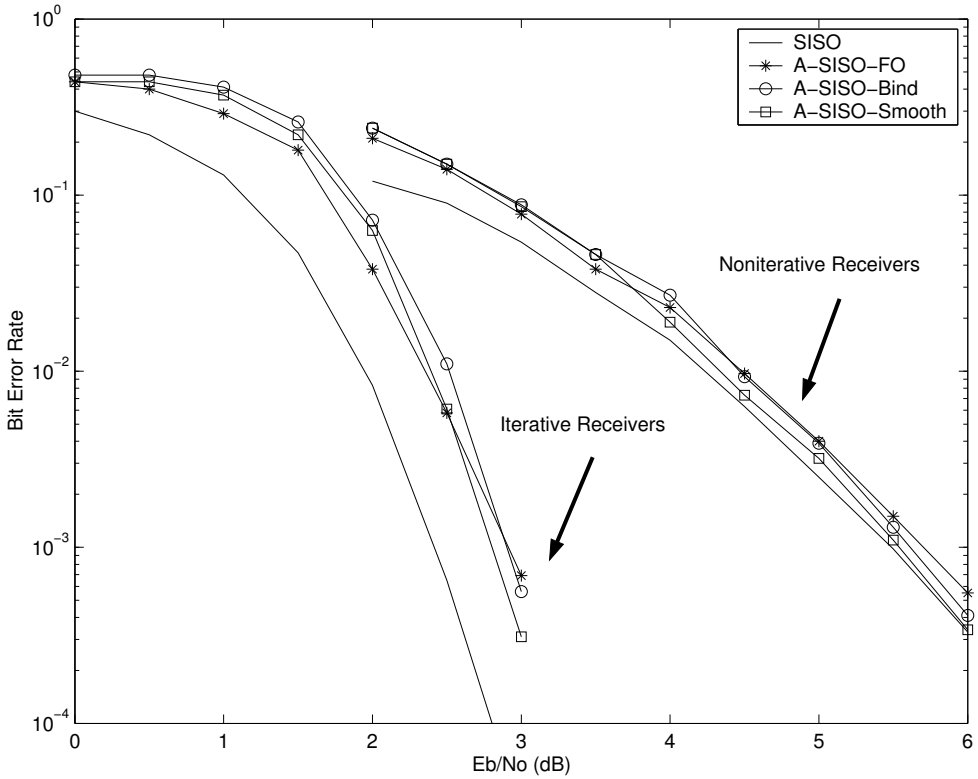


Figure 14: BER performance of the FI-A-SISO algorithms for both noniterative and iterative receivers. For iterative receivers the final decisions are made after 10 iterations. The interleaver size is $N_u = 500$.

of the SISO algorithm with ideal phase information is generated and serves as a lower bound. The same AWGN channel set in Fig. 13 is considered. At medium E_b/N_o, i.e., $E_b/N_o \geqslant 4\,\text{dB}$, the lower bound for the noniterative receiver can be closely approached by using A-SISO-Smooth. The performance gaps between A-SISO-Smooth, A-SISO-Bind and A-SISO-FO are small but still cannot be ignored. For the iterative receiver, when $E_b/N_o \geqslant 3\,\text{dB}$, using A-SISO-Smooth yields the best performance, although a fractional dB performance loss is observed compared to the lower bound. The superiority of A-SISO-Smooth in the above cases is due to the benefit of the smoothed estimation. Since E_b/N_o is high, the phase estimates in both directions are reliable and the forward and backward primary survivors are same almost everywhere. Without occurrence of phase rotations the MSK trellis acts as a normal non-RIC code and, thus, A-SISO-Smooth performs the best. However, when $E_b/N_o < 3\,\text{dB}$, A-SISO-FO outperforms A-SISO-Smooth and A-SISO-Bind for both the noniterative and iterative receivers. That is because the MSE of phase estimates is large at

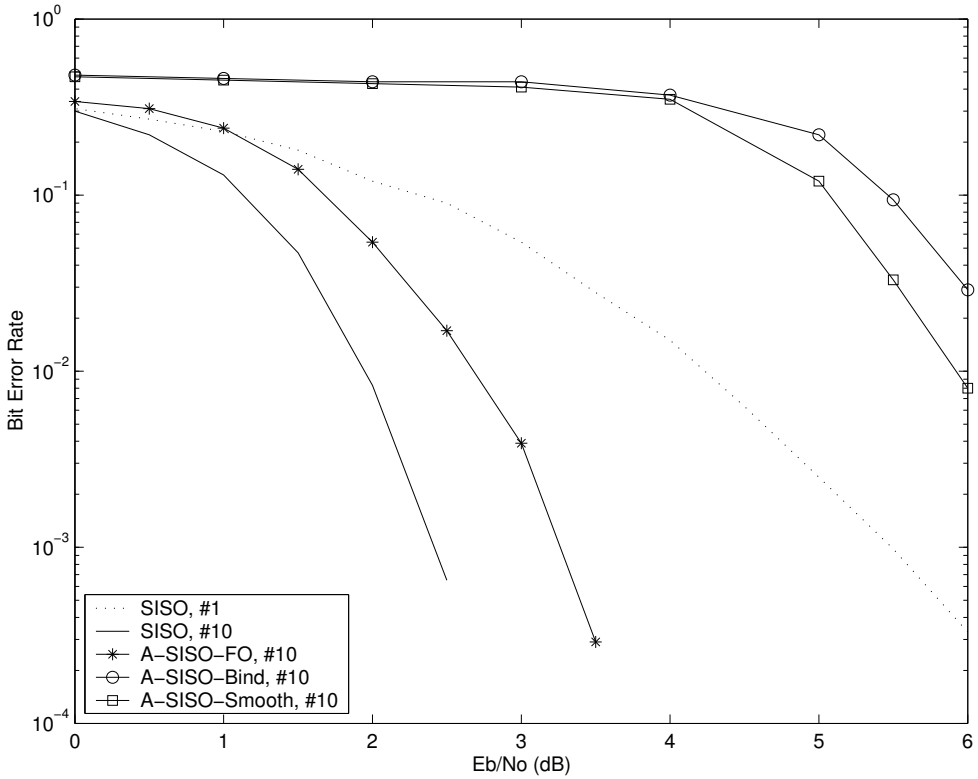


Figure 15: BER performance of the FI-A-SISO algorithms for an AWGN channel with a phase ramp. The corresponding frequency offset is $0.05/T_s$. A SCMSK system is considered. The decisions are made after 10 iterations. The interleaver size is $N_u = 500$.

low E_b/N_o and, thus, phase rotations appear. For A-SISO-Smooth and A-SISO-Bind, the forward and backward primary survivors are thus different, which causes detection errors with a high probability. For A-SISO-FO, since the primary survivors are evaluated based on the same set of phase estimates, they are identical almost everywhere, resulting in a lower BER. Based on the above demonstration, we conclude that for RIC codes, the *forward-only* FI-A-SISO algorithms are more robust than the *bidirectional* FI-A-SISO algorithms in tracking an unknown carrier phase, especially at low E_b/N_o.

To further examine this interesting phenomenon, another example is considered where an AWGN channel with an unknown phase ramp is assumed. For the first-order DPLL, a phase ramp will result in larger steady state MSE than a phase step. Therefore, the performance loss owing to the use of the *bidirectional* FI-A-SISO algorithms will be much more significant in this case. The results in Fig. 15 confirm this conjecture. The iterative receiver using A-SISO-FO exhibits at least a 3 dB power gain compared to the receiver using

A-SISO-Smooth or A-SISO-Bind.

The outcome observed in Fig. 14 and 15 is not contradictory with the conclusion stated in [42] that the *forward-only* SW-A-SISO algorithms are inferior to the *bidirectional* SW-A-SISO algorithms for the SCMSK system. Within the SW-A-SISO algorithms, joint estimation and detection for the k-th symbol interval are evaluated over the truncated observation $\mathbf{y}_{k-D}^{k+D}$, where D is the truncation depth. The challenge therein is how to initialize the backward phase estimate $\hat{\boldsymbol{\theta}}^b_{k+D|k+D+1}$ in the k-th sliding window. The usual method is to set $\hat{\boldsymbol{\theta}}^b_{k+D|k+D+1} = \hat{\boldsymbol{\theta}}_{k+D|k+D}$ [48]. Since the backward estimation recursion starts with a forward estimate, the backward and forward primary survivor would be identical with a high probability, especially when D is small. With an increase in D, there is a greater possibility that the backward primary survivor differs from the forward primary survivor, which results in a performance loss in detection.

3.4.2 Performance of the Forward-Only FI-A-SISO Algorithms

In practice, the *forward-only* FI-A-SISO algorithms are highly preferred because of their simplicity and robustness. In this subsection, they are further examined on different channels, including correlated flat fading channels as well as AWGN channels with phase jitter and a phase ramp.

A simplified version of A-SISO-FO, called A-SISO-Single, is also proposed. As shown in Appendix C, within A-SISO-FO each state is associated with a survivor and, thus, the phase estimation must be performed along multiple survivors. In A-SISO-Single, only the survivor with maximum APP is retained. In this case, the complexity reduces to $\mathcal{O}(N_u)$. The A-SISO-Single algorithm is also referred as to a SEP-based A-SISO algorithm, which is discussed in more details in Chapter IV. Moreover, we note that the innovations-based A-SISO algorithms are only one kind of A-SISO algorithms. The *forward-only* pilot-aided FI-A-SISO algorithm (A-SISO-Pilot), presented in [35], is also considered as a comparison. The value of E_b/N_o is normalized considering pilot overhead. The pilot-to-information-ratio (PIR) is a ratio between the number of pilot symbols and the number of information symbols.

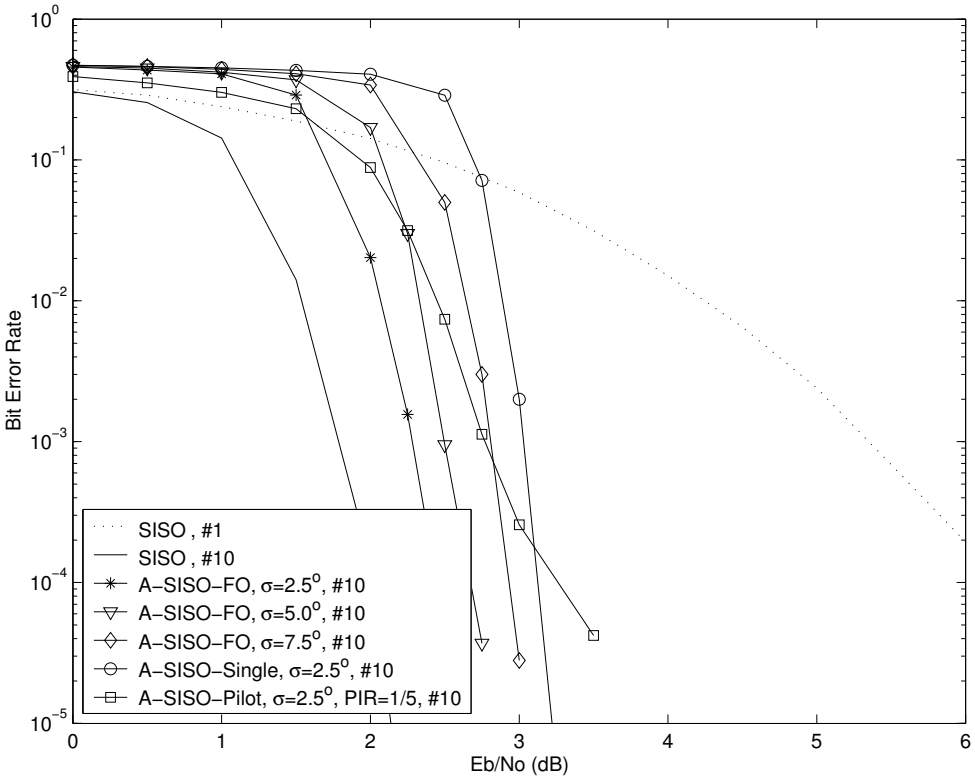


Figure 16: BER performance of the *forward-only* FI-A-SISO algorithms for an AWGN channel with different random phase jitters. A SCMSK system is considered. The decisions are made after 10 iterations. The interleaver size is $N_u = 2048$.

In Fig. 16, an AWGN channel with phase jitter is considered. Simulation results show that A-SISO-FO outperforms A-SISO-Single and A-SISO-Pilot with PIR$=1/5$. Let σ^2 be the mean square of the phase jitter, i.e., the variance of phase increment at each symbol interval. When $\sigma = 2.5°$, the complexity reduction in A-SISO-Single results in a 0.5 dB performance loss compared to A-SISO-FO, at a BER of 10^{-4}. Also, because of the poor tracking ability, A-SISO-Pilot exhibits an earlier onset of error floor at a BER of 10^{-5}, compared to A-SISO-FO and A-SISO-Single. In addition, for A-SISO-FO, the performance losses when $\sigma = 2.5°$, $5°$ and $7.5°$ are 0.5 dB, 0.7 dB and 1.0 dB, respectively, compared to the lower bound at a BER of 10^{-4}.

In Fig. 17, an AWGN channel with an unknown phase ramp is considered. The corresponding frequency offset is $0.05/T_s$. A-SISO-FO is still better than A-SISO-Single or A-SISO-Pilot with PIR$=1/5$. As for A-SISO-Pilot, at low E_b/N_o, it is totally corrupted owing to its limited tracking ability to time-variant channels. In addition, it is observed

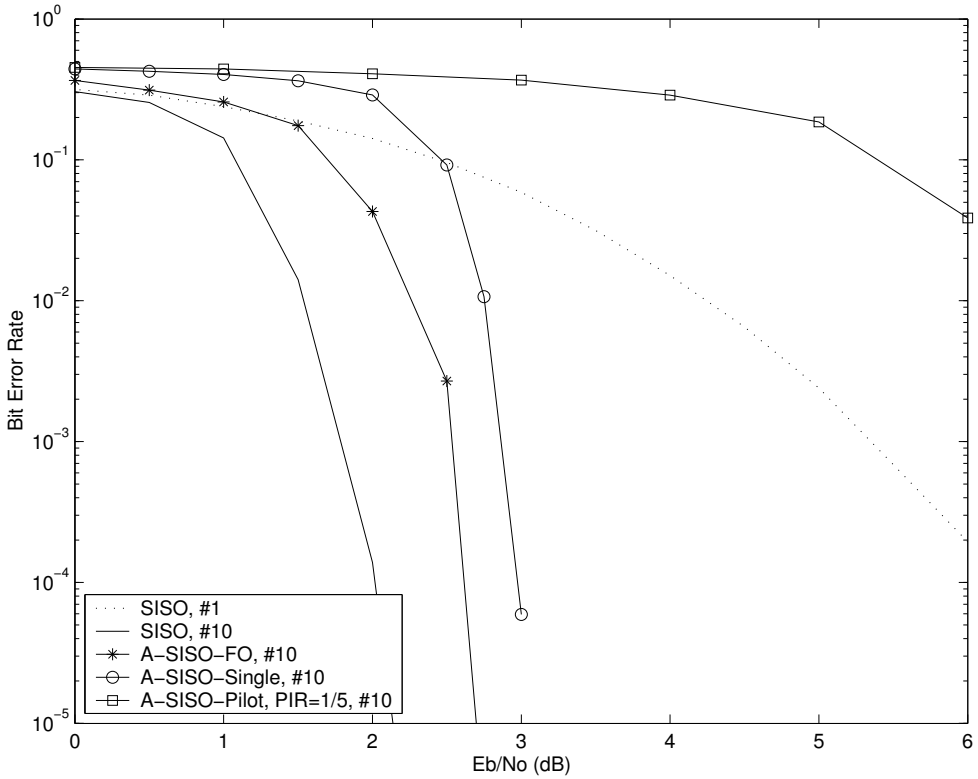


Figure 17: BER performance of the *forward-only* FI-A-SISO algorithms for an AWGN channel with a phase ramp. The corresponding frequency offset is $0.05/T_s$. A SCMSK system is considered. The decisions are made after 10 iterations. The interleaver size is $N_u = 2048$.

for A-SISO-FO that the performance loss owing to imperfect synchronization with a phase ramp is about 0.6 dB at a BER of 10^{-4}, while the performance loss is less than 0.3 dB in case of a phase step which is shown in Fig. 14.

For the correlated Rayleigh fading channel, the performance of the *forward-only* FI-A-SISO algorithms is assessed in Fig. 18. The normalized maximum Doppler shift is $f_m T_s$= 0.001. In this case, although cycle slip, false lock and hangup occur, the forward-only demodulators still work well because of rotational invariance of MSK signals. Among them, A-SISO-FO achieves the best performance, which is less than 1.5 dB away from the lower bound. Also, the performance of A-SISO-Single closely approaches that of A-SISO-FO. This implies that A-SISO-Single is a good candidate for fading channels when computational complexity of the receiver is of greater concern.

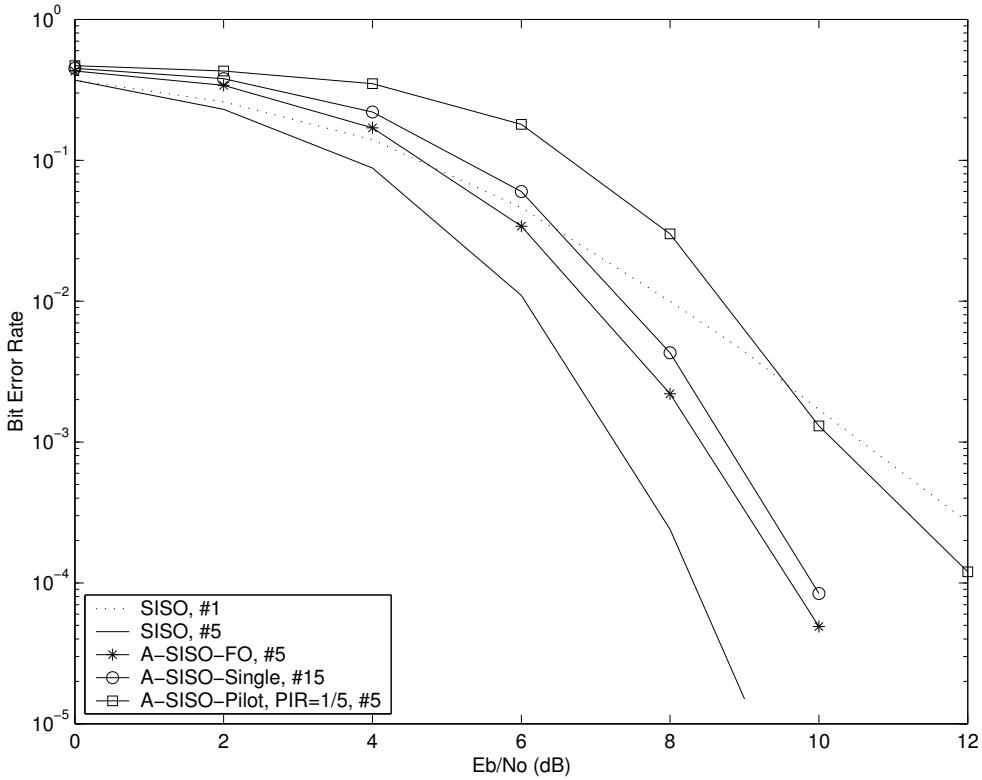


Figure 18: BER performance of the *forward-only* FI-A-SISO algorithms for a correlated Rayleigh fading channel. The normalized maximum Doppler shift is $f_m T_s = 0.001$. A SCMSK system is considered. The decisions are made after 5 iterations. The interleaver size is $N_u = 8192$.

CHAPTER IV

CONVERGENCE BEHAVIOR OF ITERATIVE PHASE SYNCHRONIZATION AND DETECTION

Serially concatenated trellis-coded modulation with rotational invariance (RI-SCTCM), including SCCPM, can tolerate certain phase ambiguities and, therefore, for these systems coherent detection is robust even when cycle slipping, false-lock and hangup exist during phase synchronization. In this chapter, we present a semi-theoretical analysis on convergence behavior of iterative phase synchronization and detection for such systems. In particular, an oscillatory convergence behavior caused by cycle slipping is explained by means of density evolution. Based on our analysis, some issues regarding to practical iterative receiver design are also discussed. The main results in this chapter can be found in [126].

4.1 Summary of Known Results

4.1.1 Serially Concatenated Trellis-coded Modulation with Rotational Invariance

A typical SCTCM transmitter is shown in Fig. 19, in which $\boldsymbol{b}$ is an information binary sequence, $\boldsymbol{c}$ and $\boldsymbol{c'}$ are the code-bit and interleaved code-bit sequences, respectively, $\boldsymbol{u}$ is the binary output at the inner encoder, and $\boldsymbol{x}$ is the modulating symbol sequence after mapping. Let R be the overall SCTCM code rate and, thus, $E_s = RE_b$, where E_s and E_b are the symbol and bit energy, respectively. For simplicity we still assume constant envelope modulation, including CPM and PSK.

A sufficient condition for a SCTCM system to be rotationally invariant is that its inner trellis-coded modulator is rotationally invariant [57]. The definition and design of RI-TCM is already summarized in 3.2.1. For RI-SCTCM, in order to achieve an interleaving gain with iterative demodulation/decoding, the inner code must be recursive. The accumulator codes (ACC) with generator polynomial $\{\frac{1}{1+D}\}$ are recursive and could be invariant to certain

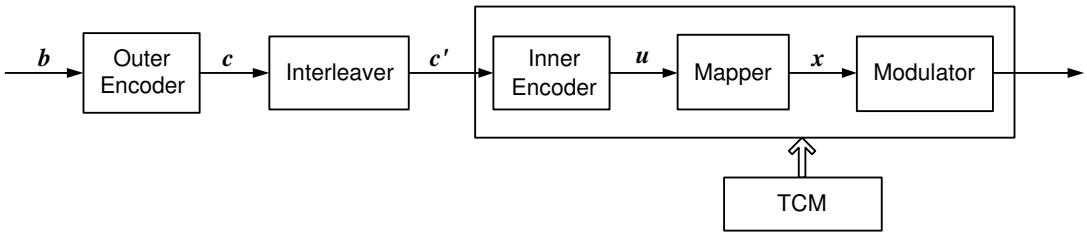
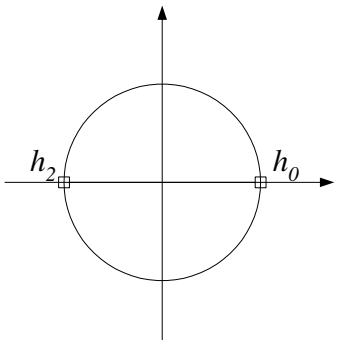


Figure 19: Transmitter structure for SCTCM.

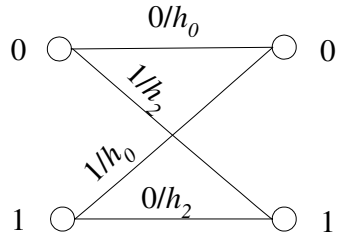
specified phase rotations. *Differential encoded coherently detected* M-ary PSK (DPSK) is a class of rate-1 ACC M-ary codes, and it is transparent to $M-1$ phase rotations including $\{\frac{2\pi}{M}, \frac{4\pi}{M}, \cdots, \frac{2\pi(M-1)}{M}\}$ [105]. Moreover, it was already shown in 3.2.2 that CPM with a rational modulation index $h_{\text{idx}} = q/p$, where q and p are relatively prime integers, is also a class of rate-$1/p$ ACC codes that is rotationally invariant to $p-1$ phase ambiguities including $\{\frac{2\pi}{p}, \frac{4\pi}{p}, \cdots, \frac{2\pi(p-1)}{p}\}$. Both DPSK and CPM are good choices as an inner code of RI-SCTCM. Fig. 20 shows the 2-state trellises of binary DPSK (DBPSK) and MSK systems. The implementation and analysis of SCDPSK and SCCPM can be found in [33] and [105].

4.1.2 Adaptive Iterative Decoding Algorithms

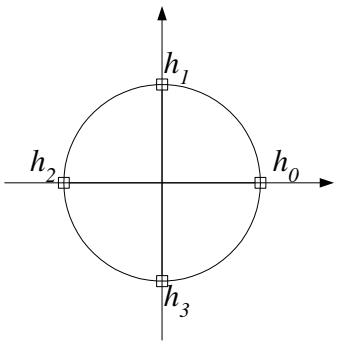
The per-survivor-processing (PSP) method was originally proposed for hard-output maximum likelihood sequence detection (MLSD) with parametric uncertainty [37]. Typically, the parameter estimation therein is embedded directly into the Viterbi algorithm, and unknown parameters are adaptively updated along with selected survivors that can be obtained by searching for paths having minimum cost. Unfortunately, the PSP principle is not directly suitable for the BCJR algorithm that is used for soft-input soft-output (SISO) MAP symbol detection (MAPSD). The BCJR algorithm searches for information symbols having the best overall statistics rather than a symbol sequence with minimum cost and, therefore, no survivor is defined and selected. To extend the PSP concept into the BCJR algorithm, survivors must be defined. In the existing PSP-based A-SISO algorithms [40], [41], the forward survivor at each state is defined as the symbol sequence with maximum forward *a-posteriori* probability among all symbol sequences ending at this state. The backward survivors are defined in a similar way. The A-SISO algorithms based on the single-estimator-processing



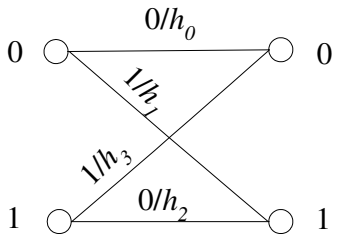
(a) DBPSK: Constellation



(b) DBPSK: Trellis



(c) MSK: Constellation



(d) MSK: Trellis

Figure 20: Two examples of RI-TCM systems

(SEP) principle can be treated as a simplified version of the PSP-based counterparts, in which all trellis states at each time epoch share a pair of forward and backward survivors.

Both PSP-based and SEP-based A-SISO algorithms can be classified into two categories, namely the forward-only and forward-backward A-SISO algorithms. Within the former, parameter estimation is processed only in the forward direction; for the latter, both forward and backward parameter estimation are required. It is shown in Chapter III that for RI-SCTCM the forward-only A-SISO algorithms are more robust than the forward-backward A-SISO algorithms in tracking a fast time-varying phase, because of rotational invariance of inner TCM. Therefore, in this chapter, we restrict our attention to the forward-only A-SISO algorithms.

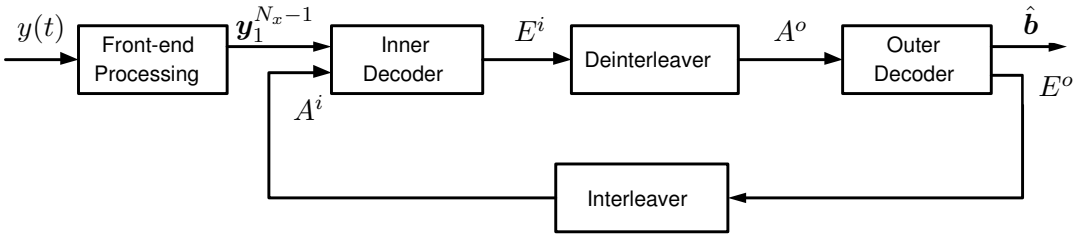


Figure 21: Iterative demodulation/decoding scheme for SCTCM.

4.1.3 Basics of Convergence Analysis

A typical iterative demodulation/decoding scheme for SCTCM is shown in Fig. 21. Convergence of iterative decoding algorithms can be analyzed by investigating exchanging of extrinsic information between constituent decoders [106],[107]. To quantify evolving reliability of decisions through iterations, mutual information is used as a measure to describe extrinsic information transfer (EXIT) characteristics [107]. Let a coded bit be a binary random variable C taking on the values $\{-1, +1\}$ with equal probability. Denote extrinsic information by a random variable E and assume that its conditional density function is symmetric, i.e., $p_E(E = \lambda|C = 1) = p_E(E = -\lambda|C = -1)$. Then the mutual information between E and C is

$$I_E(E;C) = \int_{-\infty}^{+\infty} \Big[p_E(\lambda|C=1) \times \log_2 \frac{2p_E(\lambda|C=1)}{p_E(\lambda|C=1) + p_E(\lambda|C=-1)} \Big] d\lambda. \tag{68}$$

Under the further assumptions that the extrinsic information E conditioned on C is Gaussian distributed with mean μ_E and variance σ_E^2, and it is also consistent, i.e., $2\mu_E = \sigma_E^2$, then the mutual information $I_E(E;C)$ is

$$I_E(E;C) = 1 - \mathbb{E}_{E|C=1} \left\{ \log_2(1 + e^{-E}) \right\}. \tag{69}$$

For a sufficiently large and random interleaver, histogram measurements on $p_E(E|C)$ show that the Guassian approximation and the assumption of consistency are accurate when μ_E is not large. The mutual information between the *a-priori* information A and the coded bits C can be computed in a similar way. Now we can define the extrinsic information transfer function of a SISO decoder as

$$I_E(E;C) = G(I_A(A;C), E_b/N_o), \tag{70}$$

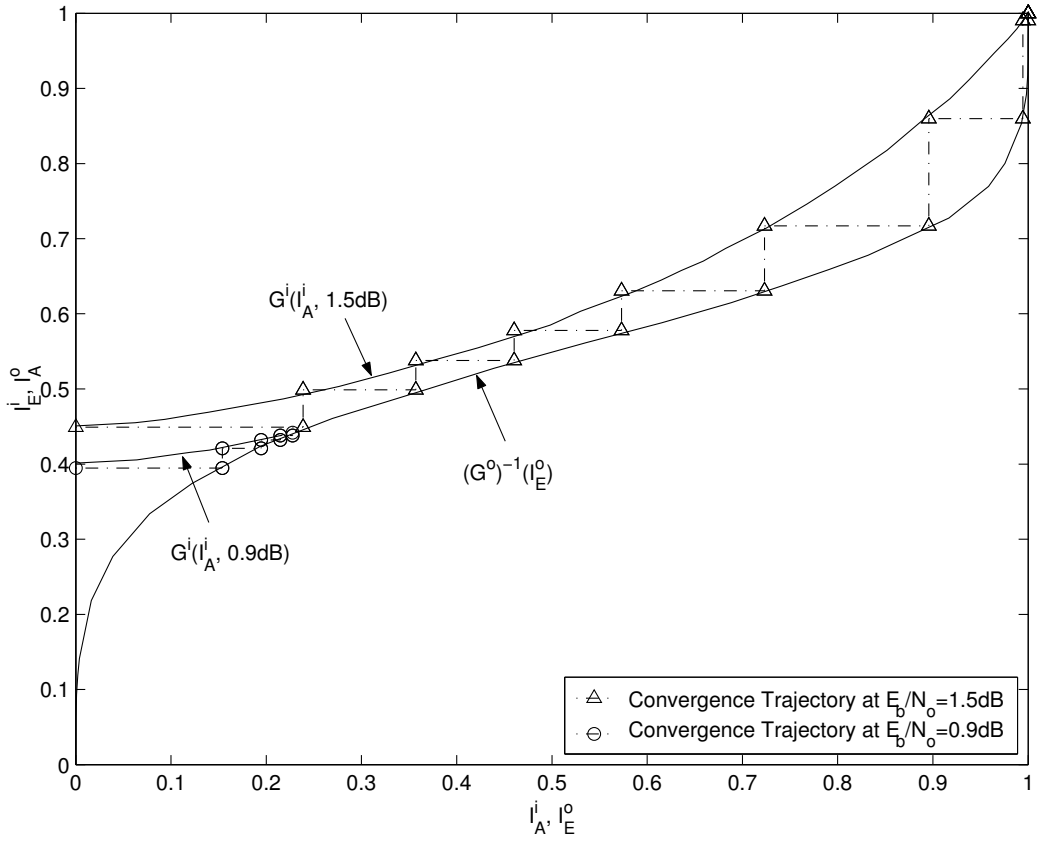


Figure 22: EXIT chart for SCMSK. $N_x = 100000$.

where the bit-energy-to-noise-density ratio E_b/N_o serves as a parameter on $G(\cdot)$. In the following we will use the superscripts "o" and "i" to distinguish between the outer and inner codes, respectively. Note that $G^o(\cdot)$ is not parameterized on E_b/N_o since no channel information is directly sent into the outer decoder.

Depicting $G^i(\cdot)$ and $(G^o)^{-1}(\cdot)$ in one diagram is referred as to the EXIT chart, in which I_A^i and I_E^o are on the abscissa, while I_E^i and I_A^o are on the ordinate. The evolution of reliability of decisions stops when the two curves intersect. Moreover, a wider separation between the two curves indicates more rapid convergence. Fig. 22 shows the convergence behavior of SCMSK with known carrier phase, where the full-complexity BCJR algorithm is used. With the outer code being a rate-1/2 convolutional code with generators $[5_8, 7_8]$, the convergence threshold is 0.9 dB. At $E_b/N_o = 1.5$ dB, the decoding trajectory sneaks through a tunnel region and converges after 8 iterations. Although not plotted in Fig. 22,

SCDBPSK has the same convergence performance as SCMSK provided they have an identical outer code. The reason is that their inner codes have the same 2-state trellis and free distance. Note that SCMSK is superior to SCDBPSK in bandwidth-efficiency. In the following analysis we will use SCMSK and SCDBPSK alternatively.

4.2 Convergence of Iterative Phase Synchronization

Analyzing the convergence of iterative phase synchronization is rather difficult and complicated for RI-SCTCM, because of existence of phase ambiguities. We introduce a simplified phase model that makes it possible to separately evaluate the effects of phase rotations and phase disturbance on convergence. Analysis on our phase model will yield valuable insights on how the PSP and SEP phase tracking methods affect the convergence of iterative receivers.

4.2.1 A Simplified Phase Model

As defined previously, the received complex envelope on a flat fading channel with additive white Gaussian noise is

$$y(t) = \sqrt{\frac{E_s}{T_s}}\, H_k \exp\{j\theta_k\} x(t, \boldsymbol{u}_1^{N_u}) + n(t), \qquad kT_s \leqslant t \leqslant (k+1)T_s, \tag{71}$$

where we assume the fading gain is fixed over each symbol interval, H_k and θ_k are the corresponding amplitude variation and phase offset owing to channel fading at the k-th symbol interval, and $n(t)$ is the AWGN with one-sided power spectral density N_o. In the simplified phase model, the phase estimate $\tilde{\theta}_{k,l}$ at the l-th state during the k-th interval is described as

$$\tilde{\theta}_{k,l} = \theta_k + \phi_{k,l} + w_{k,l}, \quad l = 0, 1, \cdots, |\mathcal{S}|-1, \phi_{k,l} \in [0, 2\pi], \tag{72}$$

where $\phi_{k,l}$ is an allowable phase ambiguity specified by the TCM structure, $w_{k,l}$ is the phase disturbance introduced by a phase estimation procedure, and $|\mathcal{S}|$ is the number of trellis states in the inner code. Any case where $\phi_{k,l} \neq 0$ indicates the occurrence of cycle slipping.

Let $\boldsymbol{w}_k = [w_{k,0}, \cdots, w_{k,|\mathcal{S}|-1}]$ be a vector of phase disturbances at the k-th time interval. The statistics of $\boldsymbol{w}_k$ depend on the type of phase synchronizer, the channel model and

the tree pruning rule within A-SISO algorithms and, therefore, they are unlikely to be determined in practice. However, in the context of least-squares or minimum mean-square-error estimation, the phase disturbance can be always viewed as a summation of a number of noise terms. From the multivariate Central Limit Theory [108], $\boldsymbol{w}_k$ can be approximated as a joint Gaussian random vector with zero mean and covariance $\boldsymbol{W}_k$. The phase estimates obtained in both PSP and SEP methods can be generalized into this phase model. For the PSP method, the elements of $\boldsymbol{w}_k$ are correlated because the parameters are estimated based on a common observation record. Within the SEP method the phase disturbance is same for all states. Finally, it is worth noting that the phase disturbance resulting from adaptive estimation is generally correlated in time.

4.2.2 Synchronization without Cycle Slipping

When the carrier phase is time-invariant or slow time-varying, the accuracy of phase tracking is usually high and cycle slips rarely occur. Also, in some applications, pilot symbols are inserted into data frames to prevent the occurrence of cycle slips [41]. More recently, non-recursive phase estimation is shown to be an efficient approach to reduce the probability of occurrence of cycle slipping [109]. In this subsection, the convergence of iterative phase synchronization without cycle slipping is investigated. According to the phase model in (72), we set $\phi_{k,l}=0$ for all possible k and l.

For clarity, we start with SCDBPSK. Let $\tilde{x}(t, \mathcal{T}(l,l'), \tilde{\theta}_{k,l})$ be the tentative detected signal over the k-th symbol interval, which is associated with the phase estimate $\tilde{\theta}_{k,l}$ and transition $\mathcal{T}(l,l')$ connecting the states l and l'. Assume that an all-zeroes sequence is transmitted over AWGN channels with $H_k = 1$ for all k. Note that the existence of random phase errors does not change the original trellis structure of an inner code. Therefore, for DBPSK, the probability of an error event with free distance from the all-zeroes path is given by [5]

$$P_{\mathrm{e,\,DBPSK}} = \mathbb{P}\Big(\|y(t) - \tilde{x}(t, \mathcal{T}(0,0), \tilde{\theta}_{k,0})\|^2 + \|y(t) - \tilde{x}(t, \mathcal{T}(0,0), \tilde{\theta}_{k+1,0})\|^2 \;\geqslant \|y(t) - \tilde{x}(t, \mathcal{T}(0,1), \tilde{\theta}_{k,0})\|^2 + \|y(t) - \tilde{x}(t, \mathcal{T}(1,0), \tilde{\theta}_{k+1,1})\|^2 \Big), \tag{73}$$

[5] We denote that $\|a(t)\|^2 = \frac{1}{T_s}\int_{T_s} a(t)a^*(t)dt$ and $\langle a(t), b(t)\rangle = \frac{1}{T_s}\int_{T_s} a(t)b^*(t)dt$.

which can be further simplified as

$$P_{\mathrm{e,\,DBPSK}} = \mathbb{P}\Big(2E_s[2\cos(w_{k,0})+\cos(w_{k+1,0})-\cos(w_{k+1,1})] \leqslant -2n_k^{(0,0)}-n_{k+1}^{(0,0)}+n_{k+1}^{(1,0)}\Big), \quad (74)$$

where $n_k^{(l,l')}=\langle n(t), \tilde{y}(t,\mathcal{T}(l,l'),\tilde{\theta}_{k,l})\rangle$ and $\mathrm{Var}(n_k^{(l,l')})=E_sN_o$. If linear parameter prediction is employed, the estimate $\tilde{\theta}_{k,l}$ is independent of channel noise $n(t)$ at the present and future time intervals. As a result, $n_k^{(0,0)}$ is independent of $n_{k+1}^{(0,0)}$ and $n_{k+1}^{(1,0)}$. Hence, we can express (74) as

$$P_{\mathrm{e,\,DBPSK}} = \frac{1}{2}\mathrm{erfc}\left(\sqrt{\frac{E_s}{N_o}\frac{[2\cos(w_{k,0})+\cos(w_{k+1,0})-\cos(w_{k+1,1})]^2}{6-2\cos(w_{k+1,0}-w_{k+1,1})}}\right). \quad (75)$$

The average error probability can be obtained by averaging on the phase disturbance. To evaluate the performance of such system, an effective E_b/N_o is defined as

$$\left(\frac{E_b}{N_o}\right)_{\mathrm{eff,\,DBPSK}} = \frac{1}{R}\Big(\mathrm{erfc}^{-1}\left(2\,\mathbb{E}_w\left\{P_{\mathrm{e,\,DBPSK}}\right\}\right)\Big)^2, \quad (76)$$

which can be computed numerically. Similarly, we can obtain the effective E_b/N_o for MSK, where the error probability is

$$P_{\mathrm{e,\,MSK}} = \frac{1}{2}\,\mathrm{erfc}\left(\sqrt{\frac{RE_b}{N_o}}\times\sqrt{\frac{[\cos(w_{k,0})+\frac{2}{\pi}\sin(w_{k,0})+\cos(w_{k+1,0})-\frac{2}{\pi}\sin(w_{k+1,1})]^2}{4+\frac{4}{\pi}\sin(w_{k+1,1}-w_{k+1,0})}}\right). \quad (77)$$

The performance loss owing to phase disturbance, defined as the difference between the effective E_b/N_o and the real E_b/N_o, is accessed in Fig. 23 for both DBPSK and MSK. For simplicity, we assume that $\boldsymbol{w}_k=[w_{k,0},w_{k,1}]$ is independent and identically distributed in time. Let the variance of phase errors at each state be σ_w^2. Denote the correlation coefficient between phase errors at two states as ρ. When $\rho=1$ the corresponding phase model represents estimates obtained by using the SEP method, while the case $-1\leqslant\rho<1$ represents the PSP method. Intuitively, the original signal constellation is alternated because of the existence of phase estimation errors and variation of ρ, which further leads to a significant change on decision region of maximum likelihood detection. For MSK, the performance loss increases with ρ and reaches the largest value when $\rho=1$. However, for DBPSK, the performance loss decreases with ρ and reaches its lowest value when $\rho=1$. This result implies that using the PSP approach, i.e., using multiple phase estimates at each symbol

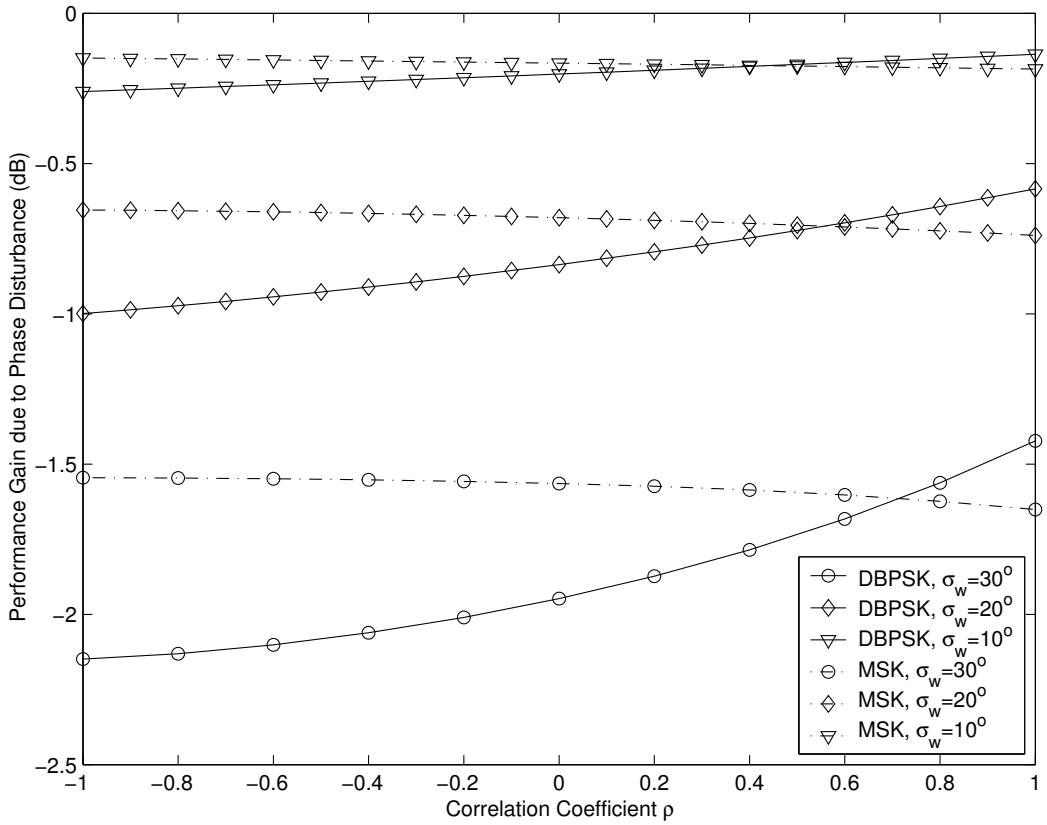


Figure 23: Performance loss owing to phase disturbance. $E_b/N_o = 3.0\,\text{dB}$

interval, does not always guarantee an improved performance. Fig. 23 also shows that both DBPSK and MSK become less sensitive to the variation of ρ when the variance of random phase errors diminishes. The performance difference caused by different values of ρ can be ignored when $\sigma_w < 10^o$. It means that the performances of PSP and SEP methods are approximately same when phase disturbance is small, which is confirmed by simulations in [41] and the section 3.4.

Fig. 24 shows the effects of phase disturbance on convergence of PSP-based and SEP-based A-SISO algorithms. The SCDBPSK system with $N_x = 100000$ operates at $E_b/N_o = 3.0\,\text{dB}$. For convenience of comparison we manipulate the phase errors in the PSP-based decoder such that $\rho = -1$ at all time stages. When $\sigma_w = 20^o$, the effective E_b/N_o are 2.42 dB and 2.00 dB for the SEP and PSP decoders, respectively, from (75)-(77). When $\sigma_w = 30^o$, the effective E_b/N_o for the SEP and PSP methods are 1.57 dB and 0.85 dB. In Fig. 24, the solid lines are the EXIT characteristics of the DBPSK SISO decoder with perfect phase

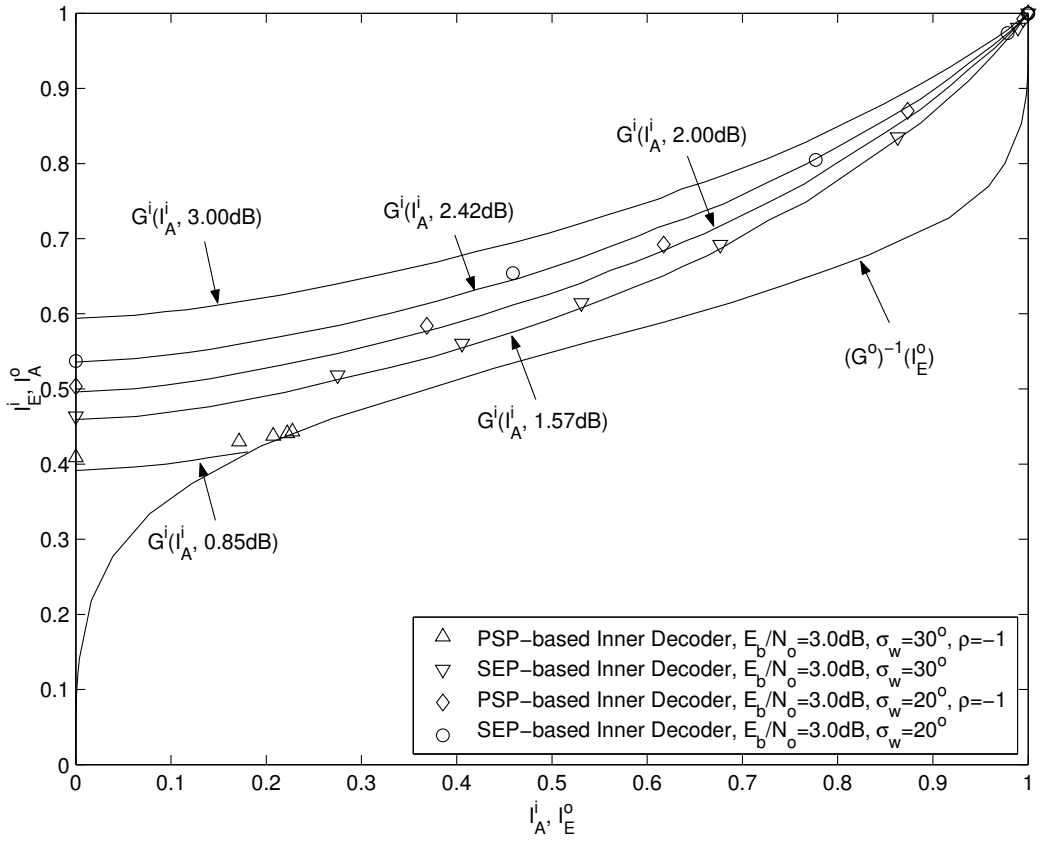
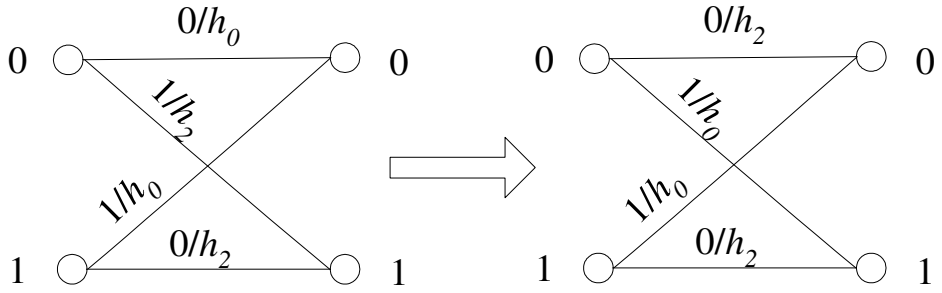


Figure 24: EXIT chart for SCDBPSK with phase disturbance. $N_x = 100000$.

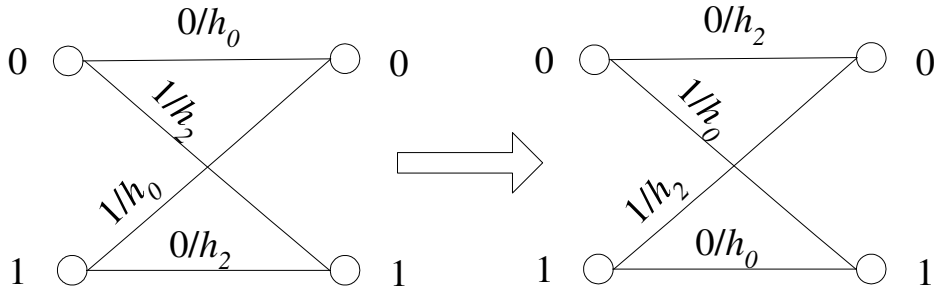
information, and generated at above effective E_b/N_o values. These curves fit well with the real convergence trajectories of iterative algorithms with corresponding phase disturbances. As expected, the SEP-based A-SISO algorithm is more robust than the PSP-based algorithm for SCDBPSK. Moreover, the results in Fig. 24 imply that without cycle slipping the effect of random phase errors can be simply viewed as an equivalent power reduction, i.e., the phase disturbance does not yield a convergence fluctuation.

4.2.3 Synchronization with Cycle Slipping

When the carrier phase has high and random dynamics, the accuracy of phase-tracking is usually low and cycle slipping frequently occurs during phase synchronization. Even when channels are time-invariant or slow time-varying, the instability of electronic devices may still cause cycle slipping. In this subsection, we will investigate the effect of cycle slipping on convergence of iterative decoding. The influence of random phase errors is ignored by



(a) Type-I cycle slipping



(b) Type-II cycle slipping

Figure 25: DBPSK trellis alternation owing to the occurrence of cycle slipping.

setting $w_{k,l}=0$ for all possible k and l. Again, DBPSK is considered as an inner recursive code.

First, we assume that cycle slipping occurs at the state 0 and the k_o-th time epoch, and phase-locking is recovered after N_{cs} symbol intervals. We also assume that the phase estimate at the state 1 keeps locked during the entire observation. Note that the allowable phase rotation for DBPSK is π and, thus, the simplified phase model becomes

$$\begin{cases} \tilde{\theta}_{k,0}=\theta_k+\phi_{k,0}, \text{ where } \phi_{k,0}=\begin{cases} \pi, \ k=k_o,\cdots,k_o+N_{cs}, \\ 0, \ \text{elsewhere;} \end{cases} \\ \tilde{\theta}_{k,1}=\theta_k, \qquad \text{for } k=0,\,1,\,\cdots,\,N_x-1. \end{cases} \tag{78}$$

We call this the Type-I cycle slipping phase model. As shown in Fig. 25 (a), the resulting trellis over $[k_oT_s,(k_o+N_{cs})T_s]$ is totally different from the original structure and, thus, the MAP detection fails on this interval. In order to correct the burst errors caused by Type-I cycle slipping, the large interleaver is required. The evolution of the *a-priori* and extrinsic

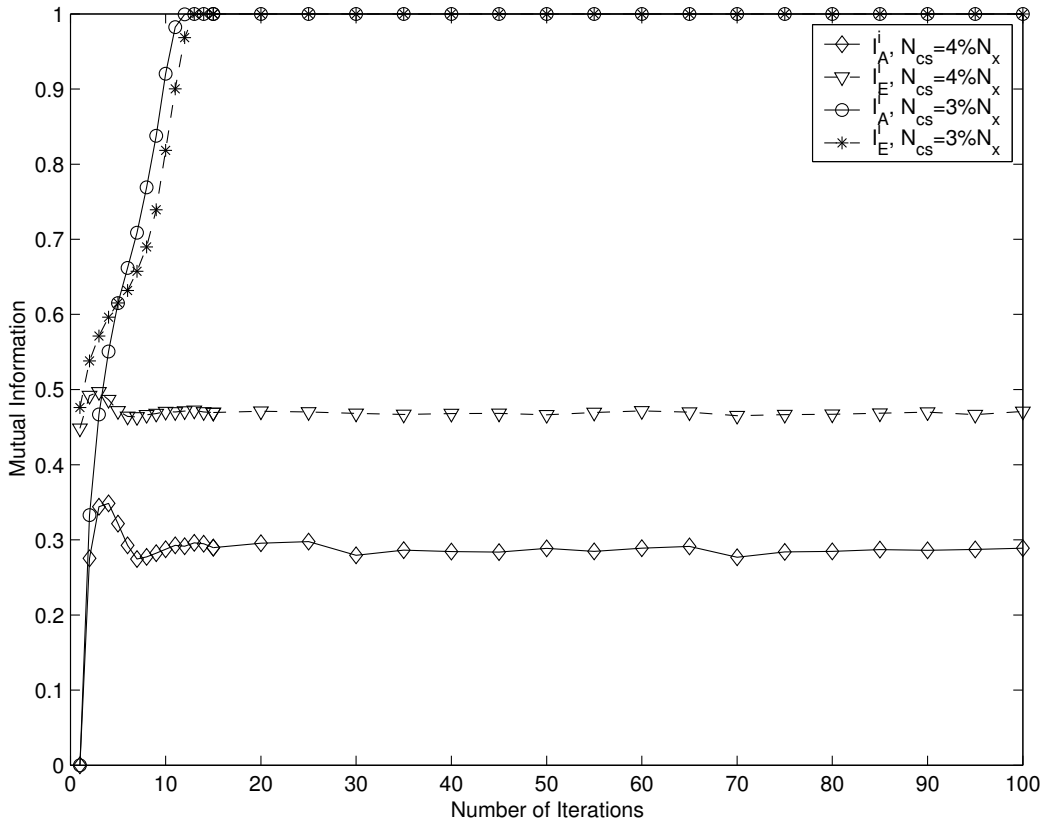


Figure 26: Evolution of mutual information at inner decoder of SCDBPSK with Type-I cycle slipping. $E_b/N_o = 3$dB. $N_x = 100000$.

information through iterations is shown in Fig. 26 for the inner decoder of SCDBPSK. At $N_x = 100000$, when the duration of false-lock is shorter than or equal to 3% of frame length, i.e., $N_{cs} \leqslant 3\%N_x$, error-free detection can be achieved by iterative decoding; otherwise, detection cannot survive the cycle slipping and a notable error floor exists. For smaller interleavers, the convergence threshold ratio between N_{cs} and N_x also becomes smaller.

The Type-II cycle slipping phase model can be described as follows:

$$\tilde{\theta}_{k,l} = \theta_k + \phi_{k,l}, \text{ where } \phi_{k,l} = \begin{cases} 0, & k = 0, \cdots, k_o - 1, \\ \pi, & k = k_o, \cdots, N_x - 1; \end{cases} \quad \text{and} \quad l = 0, 1, \tag{79}$$

where we assume that a phase rotation occurs at both states 0 and 1 at the k_o-th time epoch. The resulting trellis is shown in Fig. 25 (b), which can be obtained by flipping the original structure. Owing to the inherent rotational invariance of DBPSK, detection

does not fail when Type-II cycle slipping occurs. Moreover, if no further phase rotation occurs, the false-lock remains in effect and the phase estimates will not be driven back to the real carrier phase. The evolution of the *a-priori* and extrinsic information with decoder iterations is shown in Fig. 27. In the first few iterations, the detector converges as expected. However, after the mutual information closely approaches 1, the detector fails and afterwards convergence is gradually regained. This process repeats and results in a strange oscillatory convergence trajectory for SCDBPSK.

In order to explain this phenomenon, we refer to the standard BCJR algorithm. Denote a set of discrete observables at output of front-end processing, i.e., $\boldsymbol{y}_0^{N_x-1} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_{N_x-1}]$ as the required sufficient statistics for optimal MAPSD. At an inner decoder, the *a-priori*, extrinsic and channel information for the k-th coded bit are defined as

$$L_a(k) = \ln \frac{P(u_k = 0)}{P(u_k = 1)}, \tag{80a}$$

$$L_e(k) = \ln \frac{P(u_k = 0|\boldsymbol{y}_0^{N_x-1})}{P(u_k = 1|\boldsymbol{y}_0^{N_x-1})} - L_a(k), \tag{80b}$$

$$\text{and} \quad L_c(k) = \ln \frac{P(\boldsymbol{y}_k|x_k = h_0)}{P(\boldsymbol{y}_k|x_k = h_2)}, \tag{80c}$$

respectively. Let $\alpha_l(k)$ be the k-th forward state metric at the state l. For the 2-state DBPSK trellis, the forward metric recursion with ideal knowledge of the carrier phase is

$$\alpha_0(k+1) = \alpha_0(k)P(u_k = 0)P(\boldsymbol{y}_k|x_k = h_0) + \alpha_1(k)P(u_k = 1)P(\boldsymbol{y}_k|x_k = h_0), \tag{81a}$$

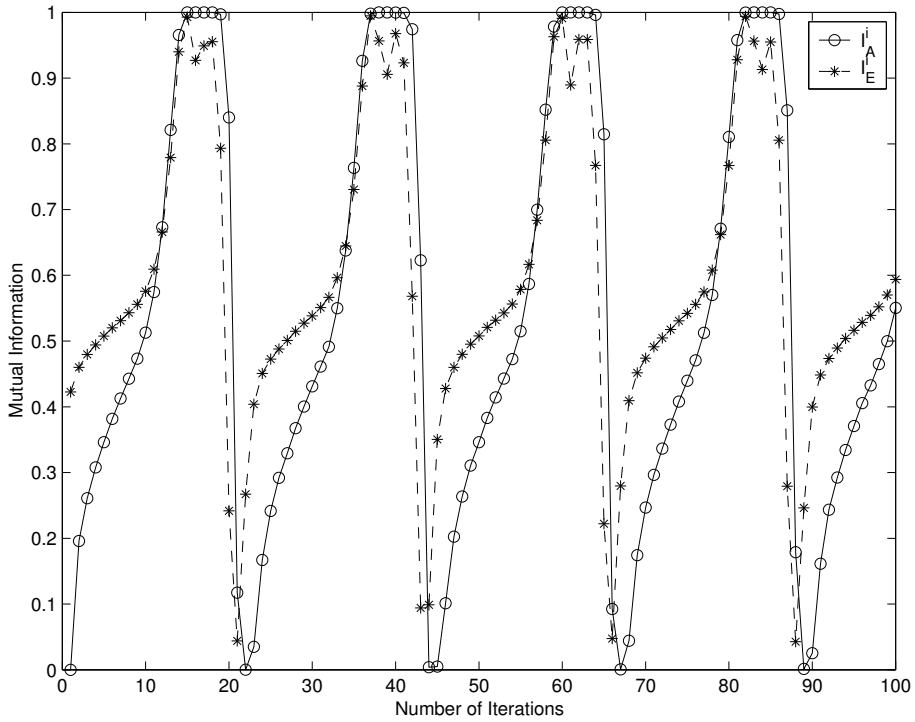
$$\alpha_1(k+1) = \alpha_0(k)P(u_k = 1)P(\boldsymbol{y}_k|x_k = h_2) + \alpha_1(k)P(u_k = 0)P(\boldsymbol{y}_k|x_k = h_2). \tag{81b}$$

By introducing the log-ratio of forward state metrics, the forward recursion can be expressed in a more compact form as follows [110]:

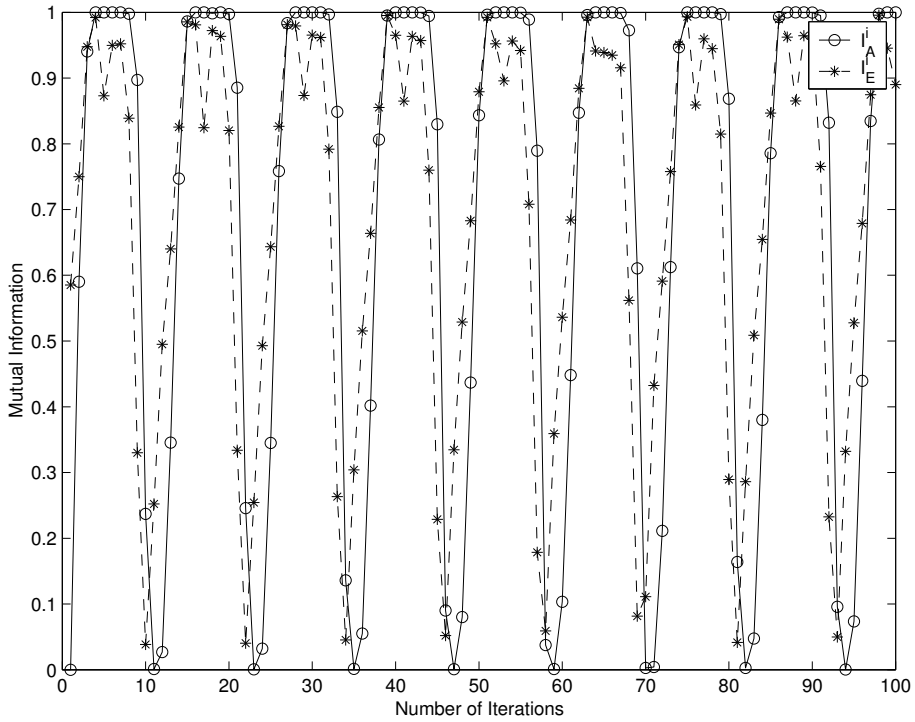
$$L_\alpha(k+1) = \ln \frac{\alpha_0(k+1)}{\alpha_1(k+1)} = L_c(k) + \ln \frac{e^{L_\alpha(k)+L_a(k)}+1}{e^{L_\alpha(k)}+e^{L_a(k)}}. \tag{82}$$

We can define and approximate the second term in (82) as

$$\begin{aligned} \min^*(L_\alpha(k), L_a(k)) &= \ln \frac{e^{L_\alpha(k)+L_a(k)}+1}{e^{L_\alpha(k)} + e^{L_a(k)}} \\ &\approx \operatorname{sign}(L_\alpha(k))\operatorname{sign}(L_a(k)) \min(|L_\alpha(k)|, |L_a(k)|). \end{aligned} \tag{83}$$



(a) $E_b/N_o = 1.2\,\mathrm{dB}$



(b) $E_b/N_o = 3.0\,\mathrm{dB}$

Figure 27: Evolution of mutual information at inner decoder of SCDBPSK with Type-II cycle slipping. $N_x = 100000$

Similarly, the backward recursion and extrinsic information can be expressed as

$$L_\beta(k) = \min^*(L_\beta(k+1)+L_c(k),\, L_a(k)), \tag{84}$$

$$L_e(k) = \min^*(L_\alpha(k),\, L_\beta(k+1)+L_c(k)), \tag{85}$$

where $L_\beta(k)$ is the log-ratio of backward state metrics at the k-th time stage. Under the assumption that an all-zeroes sequence is transmitted, the mean of $L_a(k)$, $L_e(k)$, $L_\alpha(k)$ and $L_\beta(k)$ are all positive and grow gradually through iterations according to the concept of density evolution [106]. But the mean of $L_c(k)$ is fixed at all iterations and proportional to E_b/N_o.

When Type-II cycle slipping occurs, the forward metric recursion after the k_o-th time stage becomes

$$\alpha_0(k+1) = \alpha_0(k)P(u_k=0)P(\boldsymbol{y}_k|x_k=h_2) + \alpha_1(k)P(u_k=1)P(\boldsymbol{y}_k|x_k=h_2), \tag{86a}$$

$$\alpha_1(k+1) = \alpha_0(k)P(u_k=1)P(\boldsymbol{y}_k|x_k=h_0) + \alpha_1(k)P(u_k=0)P(\boldsymbol{y}_k|x_k=h_0). \tag{86b}$$

As a result, its compact form in terms of log-ratio is,

$$L_\alpha(k+1) = -L_c(k) + \min^*(L_\alpha(k),\, L_a(k)). \tag{87}$$

Similarly, the backward metric recursion and extrinsic information become

$$L_\beta(k) = \min^*(L_\beta(k+1)-L_c(k),\, L_a(k)), \tag{88}$$

$$L_e(k) = \min^*(L_\alpha(k),\, L_\beta(k+1)-L_c(k)). \tag{89}$$

Since the trellis flips after the k_o-th time epoch, the mean of $L_a(k)$ is positive when $k \leqslant k_o$ and the mean of $L_\beta(k)$ is negative when $k \geqslant k_o+1$. In order to generate reliable extrinsic information at $k \geqslant k_o+1$, it is clear from (89) that $L_\alpha(k)$ and $L_\beta(k+1)-L_c(k)$ must have the same sign. Therefore, for $k \geqslant k_o+1$, $L_\alpha(k)$ usually must be negative to guarantee correct detection. Although $L_\alpha(k_o)$ is usually positive, the mean of the channel information becomes negative after the k_o-th time stage, which guarantees that $L_\alpha(k)$ will gradually become negative after a finite duration. The length of this duration, $N_{\alpha,\mathrm{err}}$, is roughly approximated as

$$N_{\alpha,\,\mathrm{err}} \approx \left\lceil \frac{\min(L_\alpha(k_o), \mathbb{E}\{L_a(k)\})}{\mathbb{E}\{L_c(k)\}} \right\rceil, \tag{90}$$

by assuming that $L_c(k)$ and $L_a(k)$ are fixed over this duration and equal to their respective means. In (90), $\lceil \cdot \rceil$ represents the ceil function. Similarly, for $k \leqslant k_o$, a period is required to drive $L_\beta(k)+L_c(k)$ back to a positive value. The length of this period is given by

$$N_{\beta,\,\mathrm{err}} \approx \left\lceil \frac{\min(|L_\beta(k_o)|, \mathbb{E}\{L_a(k)\})}{\mathbb{E}\{L_c(k)\}} \right\rceil. \tag{91}$$

We call the duration $[(k_o - N_{\beta,\,\mathrm{err}})T_s, (k_o + N_{\alpha,\,\mathrm{err}})T_s]$ the *error neighborhood* of cycle slipping since during this period the extrinsic information is unreliable.

Now let us return to the oscillatory convergence in Fig. 27. At the first few iterations, the reliability of the *a-priori* information is low and, thus the *error neighborhood* is narrow from (90) and (91). Although the extrinsic information at output of inner decoder has errors, thanks to the interleaver and outer decoder the reliability of extrinsic information at output of outer decoder keeps improving. As a result, the mean of the *a-priori* information at input of inner decoder also increases, and so does the size of the *error neighborhood*. Once the effect of unreliable extrinsic information resulting from the *error neighborhood* exceeds the error-correcting capability of outer code, the detector will consequently fail. Afterward, the mean of the *a-priori* information at inner decoder drops back to a low value in the following one or two iterations. From then on, the effect of the *error neighborhood* becomes less dominant and, thus, convergence is regained. This procedure repeats and results in a convergence fluctuation in Fig. 27. At a higher E_b/N_o, it usually takes less iterations for the *a priori* information to converge, yielding an earlier onset of convergence oscillations.

Another explanation for the oscillatory convergence behavior can be drawn based on the reformulation of the distribution of extrinsic information. Recall that with perfect synchronization the conditional density function of extrinsic information $p_E(E|C)$ is Gaussian and consistent. With Type-II cycle slipping these assumptions are still valid when the conditional mean μ_E is low. However, when μ_E is large, $p_E(E|C)$ is neither Gaussian nor consistent. Experimental results show that $p_E(E|C)$ at the inner decoder can be approximated as

$$p_E(E = \lambda | C = +1) = (1 - 2\kappa\mu_g)\mathcal{N}(\mu_g, \sigma_g^2) + \kappa\mu_g\,\mathcal{H}(-\mu_g, +\mu_g), \tag{92}$$

where $\mathcal{N}(\mu_g, \sigma_g^2)$ is the density function of a Gaussian random variable with mean μ_g and

variance σ_g^2, $\mathcal{H}(-\mu_g, +\mu_g)$ is the density function of a random variable uniformly distributed over $[-\mu_g, +\mu_g]$, and κ is a constant depending on the structure of inner trellis and the size of interleaver. Intuitively, the first term in (92) corresponds to reliable detection outside the *error neighborhood*, while the second term in (92) is associated with the conditional density of extrinsic information generated inside the *error neighborhood*. Although κ is usually very small, e.g., $\kappa \approx 10^{-4}$ for DBPSK with $N_x = 100000$, the error tail of $p_E(E = \lambda | C = +1)$ on the region of $(-\mu_g, 0]$ becomes notable when μ_g is large. As μ_g grows through iterations, the effect of this tail keeps being enforced until the detection fails. Then, μ_g drops back, the effect of the tail diminishes and convergence is regained. This procedure repeats as shown in Fig. 27.

In summary, both Type-I and Type-II cycle slipping can significantly degrade the convergence performance of RI-SCTCM, although in different ways. For a real iterative phase synchronizer using the PSP approach, there usually exist both Type-I and Type-II cycle slipping and, thus, both an error floor and a convergence oscillation might appear. For the practical SEP-based A-SISO decoders, only Type-II cycle slipping occurs since all states possess an identical parameter estimate. Therefore, in the SEP method the convergence of joint synchronization and demodulation is always guaranteed if a proper iteration-stopping criterion is used.

4.3 The Effect of Doppler Spread on Mean Time to Cycle Slipping

Cycle slipping plays a crucial role on convergence of iterative phase synchronization. The statistics of cycle slipping are therefore of interest. Among them, the mean time to the first cycle slipping is an important characteristic and can be evaluated by solving the Fokker-Planck equation [9]. However, solving this equation is not trivial even for the first-order phase-locked loops (PLL) with noise-induced cycle slipping [111]. It is more difficult to solve it in the presence of channel fading. In this section, we propose a numerical approach to evaluate the effect of Doppler on the mean time to cycle slipping.

A general synchronizer model is shown in Fig. 28. A continuous-time channel model

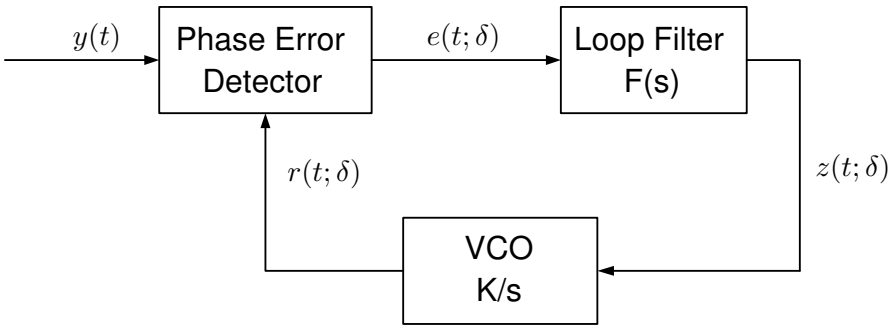


Figure 28: Structure of phase-locked loops.

with flat fading and AWGN is assumed in this section. The local reference $r(t)$ is generated by a voltage controlled oscillator (VCO), and the signal $e(t;\delta)$ contains the phase error $\delta(t)$. The loop filter is a low-pass filter with transfer function $F(s)$. Its output $z(t;\delta)$ can be decomposed as

$$z(t;\delta) = \mathbb{E}_{n_{LP}}\{z(t;\delta)\} + n_{LP}(t;\delta), \tag{93}$$

where $n_{LP}(t;\delta)$ is the zero mean low-pass filtered phase noise with power spectral density $G_{LP}(\delta)$ [6]. Note that $n_{LP}(t;\delta)$ does not need to be Gaussian. Here we only consider a first-order loop with a sinusoid phase detector and, thus, the frequency of the VCO is determined by

$$\frac{d\delta}{dt} = K_2 K_1 \sin(\delta) + K_2 n_{LP}(t;\delta), \tag{94}$$

where K_2 is the gain of the VCO, and $K_1 \sin(\delta) = \mathbb{E}_{n_{LP}}\{z(t;\delta)\}$. Equation (94) is a typical Fokker-Planck equation and its solution is a Markov process. We assume that the initial phase error is $\delta(0)=0$. Let the average time for $\delta(t)$ to reach -2π or 2π be the mean time to cycle slipping, T_{cs}. As documented in [111], the formal expression for T_{cs} is

$$T_{cs} = \int_{-2\pi}^{2\pi} \frac{2e^{-U(\delta)}}{K_{00}(\delta)} \int_{-2\pi}^{\delta} [D_0 - U_0(\zeta)] e^{U(\zeta)} d\zeta \, d\delta, \tag{95}$$

where $U_0(\cdot)$ is the unit step function, D_0 is defined as

$$D_0 = \frac{\int_0^{2\pi} e^{U(\delta)} d\delta}{\int_{-2\pi}^{2\pi} e^{U(\delta)} d\delta}, \tag{96}$$

[6] In PLLs, the bandwidth of loop filter is usually much smaller than the symbol rate. As a result, at the vicinity of $f = 0$, the power spectral density of phase noise can be approximated as $G_{LP}(f;\delta) \approx G_{LP}(0;\delta) = G_{LP}(\delta)$.

and the potential function $U(\delta)$ is given by

$$U(\delta) = \int_{\min(\delta,0)}^{\max(\delta,0)} \frac{-2K_0(\zeta)}{K_{00}(\zeta)} d\zeta. \tag{97}$$

The functions $K_0(\delta)$ and $K_{00}(\delta)$ are called intensity coefficients and given by

$$K_0(\delta) = -K_2 K_1 \sin(\delta) + \frac{1}{4} K_2^2 G'_{LP}(\delta), \tag{98a}$$

$$K_{00}(\delta) = K_2^2 G_{LP}(\delta), \tag{98b}$$

where $G'_{LP}(\delta)$ is the derivative of $G_{LP}(\delta)$.

Determining $G_{LP}(\delta)$ is critical for finding the mean time to cycle slipping. The exact derivation of $G_{LP}(\delta)$ is not available in the literature even for AWGN channels. Here we present a simple method to approximate $G_{LP}(\delta)$ for both AWGN and time-selective fading channels. First, for simplicity, we disregard the phase-error dependence of $G_{LP}(\delta)$ by assuming that $G_{LP}(\delta)$ is a constant. Then, we approximate the low-pass filtered phase noise as

$$n_{LP}(t) = z(t;\delta) - \lim_{m \to +\infty} \frac{1}{mT_s} \int_0^{mT_s} z(t;\delta) dt. \tag{99}$$

which is valid under the assumption that $z(t;\delta)$ is ergodic. Finally, we obtain

$$G_{LP} \approx \lim_{m \to +\infty} \frac{1}{mT_s} \int_0^{mT_s} n_{LP}^2(t) dt. \tag{100}$$

The normalized mean time to cycle slipping over AWGN and flat fading channels is shown in Fig. 29, in which the power spectral density of $n_{LP}(t)$ is approximately computed by using (99) and (100). Also we assume that the loop bandwidth B_L is fixed for convenience of comparison. It is shown in Fig. 29 that the presence of fading reduces the average time to cycle slipping, and the greater Doppler spread, the less mean time required for cycle slipping. Although the normalized mean time shown in Fig. 29 is rather optimistic compared practical values (see [112] for example), it gives us the insights on how fading affects the robustness of phase tracking.

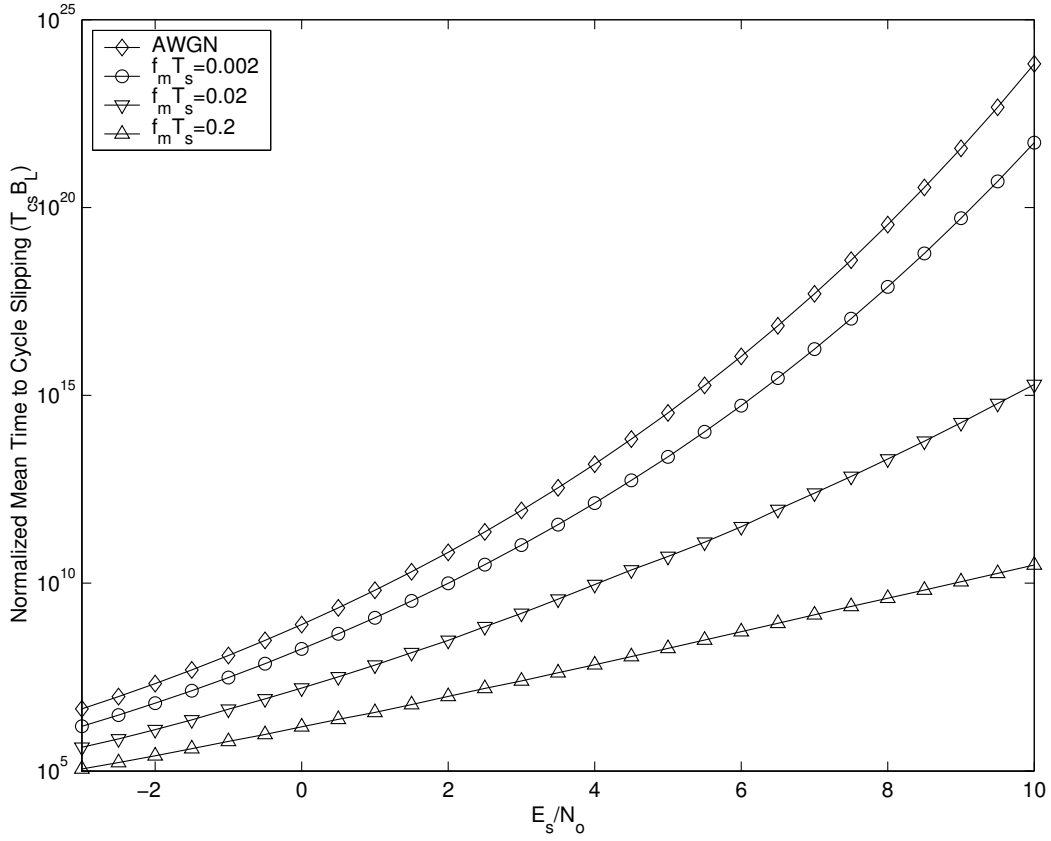


Figure 29: Mean time to cycle slipping over AWGN and time-selective fading channels.

4.4 Numerical Results And Discussions

The performance of iterative phase synchronization without the aid of pilot symbols has been evaluated by extensive simulations for SCCPM systems. The design issues, including iteration-stopping criteria, interleaver size, and decision depth of SEP-based A-SISO algorithms, are discussed. At the transmitter, a pseudo-random interleaver and a rate-1/2 convolutional outer code with generators $[5_8, 7_8]$ are employed. At the receiver, a first-order digital PLL is considered within A-SISO demodulation, and its closed-loop gain is optimized such that BER is always minimized. Moreover, we assume for simplicity that frequency synchronization, timing recovery and noise variance estimation have been perfectly achieved at the receiver.

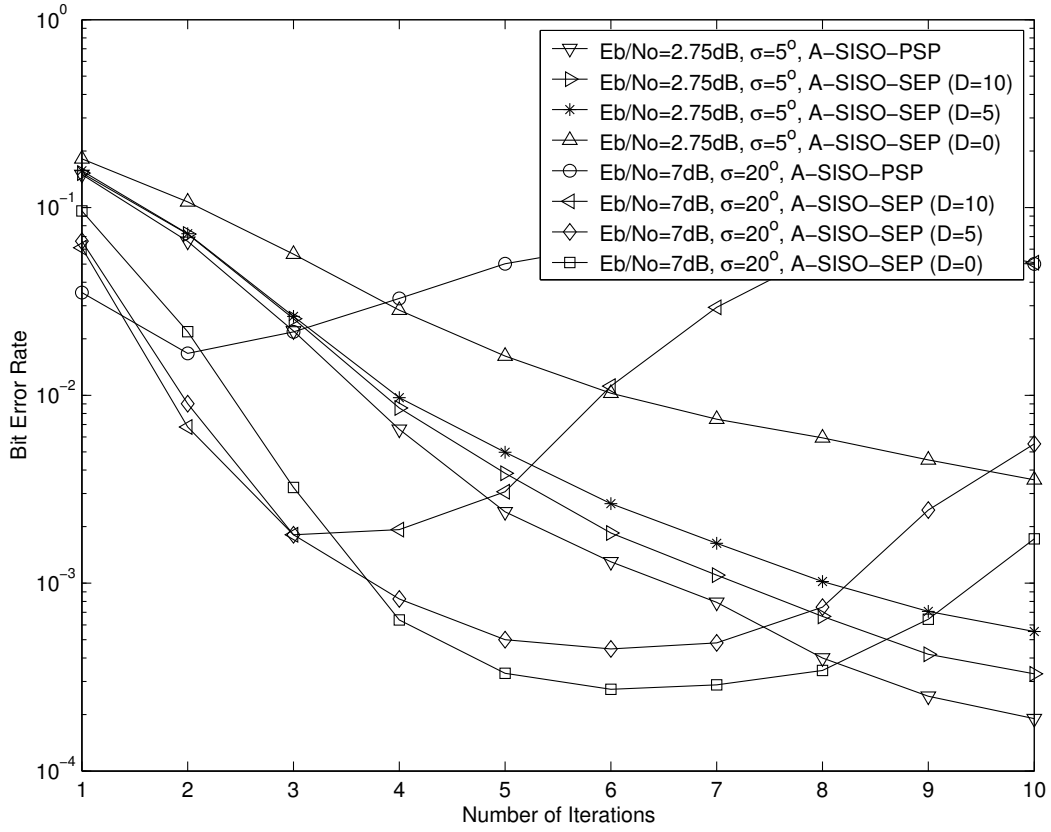


Figure 30: Effect of decision depth on the SEP-based A-SISO algorithms. SCMSK is considered with $N_x = 2048$.

4.4.1 Effect Of Decision Depth for SEP-based A-SISO Algorithms

In the PSP method, no trace-back operation is required in selecting survivors, and parameter estimates are processed in a decision-directed manner with zero delay. However, in the SEP method, only one tentative decision is retained at each symbol interval and, therefore, a trace-back operation is normally needed to increase decision reliability. For example, in the SEP-based A-SISO algorithms [41], the k-th tentative decision is usually generated by a backward recursion with decision depth D. Since no present and future parameter estimates are available at the k-th time epoch, the $(k-1)$-th estimate is used in this D-lag backward metric recursion running from the $(k+D)$-th to the k-th time stage. The complexity of estimation procedure with the D-lag SEP method is on order of $\mathcal{O}((D+1)N_x)$, while the complexity of estimation in the PSP method is on order of $\mathcal{O}(N_s N_x)$, where N_s is the number of survivors retained in the PSP method.

The BER performance of SEP-based A-SISO algorithms with different decision depths is shown in Fig. 30. The SCMSK system is considered over an AWGN channel with random phase jitters that can be modelled as a standard Wiener process. When the channel has low dynamics, e.g., the root-mean-square (rms) of phase jitters is $\sigma = 5^o$, the accurate phase-tracking can be achieved and usually no cycle slipping occurs. As a result, both SEP-based and PSP-based A-SISO demodulator consistently converge through iterations. Moreover, owing to the slow time-varying property of the channel, using obsolete estimates in the trace-back recursion does not impair the decision reliability. As expected, it is found that the BER performance of SEP-based A-SISO demodulators keeps improving with increasing D, and can closely approach that of PSP-based A-SISO demodulators when D is large enough.

However, when a channel has high dynamics, e.g., the rms of phase jitters is $\sigma = 20^o$, the accurate phase-tracking becomes impossible and, thus, cycle slipping occurs. In the PSP-based A-SISO demodulator, both Type-I and Type-II cycle slips occur with high probability. Therefore, the PSP-based demodulator exhibits a notable error floor after an early convergence. For the SEP-based demodulator, only Type-II cycle slips can occur, and, thus, an oscillatory convergence behavior is observed. Since the channel is fast time-varying, employing obsolete estimates in the D-lag backward recursion reduces the reliability of tentative decisions, and further degrades the robustness of future estimation and detection. It is shown in Fig. 30 that the greater decision depth results in an earlier convergence fluctuation and a higher error floor. The SEP-based demodulator with $D = 0$ is shown to be more robust and less complex.

4.4.2 Selection of Iteration-stopping Criterion

As shown in Fig. 30, the convergence fluctuation is observed at both PSP-based and SEP-based demodulators. An intuitive solution toward this end is to terminate iterative receivers once detection starts to fluctuate. In [59], the average entropy of information bits is proposed as a useful stopping criterion, in which the iteration is terminated once the average entropy

starts to increase. The average entropy of information bits can be calculated as,

$$\tilde{H}(\boldsymbol{b}) = \frac{1}{N_b}\sum_{k=0}^{N_b-1}\sum_{m=0}^{1} -[P(b_k=m|\boldsymbol{r}_0^{N_x-1})\ln P(b_k=m|\boldsymbol{r}_0^{N_x-1}), \tag{101a}$$

$$= \frac{1}{N_b}\sum_{k=0}^{N_b-1}\left[\ln(1+e^{L_b(k)}) - \frac{e^{L_b(k)}}{e^{L_b(k)}+1}L_b(k)\right], \tag{101b}$$

where N_b is the length of the information bit sequence, and $L_b(k)$ is the log *a-posteriori* probability ratio of the k-th information bit and defined as,

$$L_b(k) = \ln\frac{P(b_k=0|\boldsymbol{r}_0^{N_x-1})}{P(b_k=1|\boldsymbol{r}_0^{N_x-1})}. \tag{102}$$

To guarantee computational stability, the equation (101b) rather than (101a) is used in practice.

We propose a novel iteration-stopping criterion, based on the concept of density evolution. Recall that extrinsic information at output of outer decoder can be treated as a symmetric random variable E^o with conditional mean μ_{E^o} and conditional variance $\sigma^2_{E^o}$. In [106], an equivalent signal-to-noise ratio (*SNR*) of this random variable is defined as *SNR* $= \mu^2_{E^o}/\sigma^2_{E^o}$, not to be confuse with E_b/N_o. Typically, a higher value of *SNR* corresponds to more reliable extrinsic information. As indicated in Section 4.2, the occurrence of cycle slips results in a large tail in the conditional density function of E^o, which further leads to an increase in $\sigma^2_{E^o}$ and a corresponding decrease in *SNR*. Therefore, *SNR* could serve as a good stopping measure for iterative phase synchronization and demodulation, in which the detection procedure is terminated once *SNR* starts to reduce. Since the transmitted code bits are *a-priori* unknown at the receiver, we simply approximate μ_{E^o} and $\sigma^2_{E^o}$ at each iteration as

$$\tilde{\mu}_{E^o} = \frac{1}{N_c}\sum_{k=0}^{N_c-1}|L_e^o(k)|, \tag{103a}$$

$$\tilde{\sigma}^2_{E^o} = \frac{1}{N_c}\sum_{k=0}^{N_c-1}(|L_e^o(k)| - \tilde{\mu}_{E^o})^2, \tag{103b}$$

where $L_e^o(k)$ is the extrinsic information for the k-th code bit at output of outer decoder.

Table 3: Performances of different iteration-stopping criteria.[a]

Stopping Criteria $\Longrightarrow$		Ideal	*SNR*	Entropy	Entropy-L	Fixed-Iter.
A-SISO-PSP	BER	8.0×10^{-4}	1.2×10^{-3}	5.2×10^{-3}	5.5×10^{-3}	3.2×10^{-3}
	Avg. Iters.	2.50	2.67	3.36	4.36	1.00
A-SISO-SEP	BER	1.1×10^{-4}	3.6×10^{-4}	4.8×10^{-4}	1.1×10^{-3}	2.6×10^{-4}
(D=0)	Avg. Iters.	2.9	6.12	5.88	6.88	2.00

[a] SCMSK operates on a time-selective fading channel with $f_m T_s = 0.01$. $E_b/N_o = 10\,\text{dB}$. $N_x = 2048$.

The performances of receivers using different iteration-stopping criteria are compared in Table 3. The SCMSK system is considered over a fading channel with normalized Doppler frequency $f_m T_s = 0.01$. Besides the above stopping criteria based on the average entropy and *SNR*, two other criteria are also considered for comparison. One is called the ideal iteration-stopping criterion, in which the receiver stops once the number of decision errors increases. Since information bits are *a-priori* unknown at the receiver, this criterion is actually unlikely to be realized in practice. Another one is called the fixed-iteration stopping criterion, in which turbo detection stops after a fixed number of iterations whenever convergence has been achieved. With the current setup, the PSP-based detector achieves its minimum BER at the first iteration, while the SEP-based detector at the second iteration. However, the number of iterations required for the minimum BER varies with system configuration and channel environments and, thus, it is impossible to predetermine it for a real receiver. For the average-entropy, *SNR* or ideal stopping criterion, when detection is terminated, the data decisions at the previous iteration, rather than at the present iteration, will be output as final decisions. The reason is that the detection has already become unreliable when it stops. The performance of the average-entropy criterion with final decisions generated at the last iteration [59], denoted as Entropy-L, is also included as a comparison in Table 3. It is shown that in terms of BER, the receiver using the *SNR* stopping criterion is superior to the receiver using the average-entropy and fixed-iteration criteria, and slightly worse than the receiver using the ideal stopping criterion. In terms of average number of required

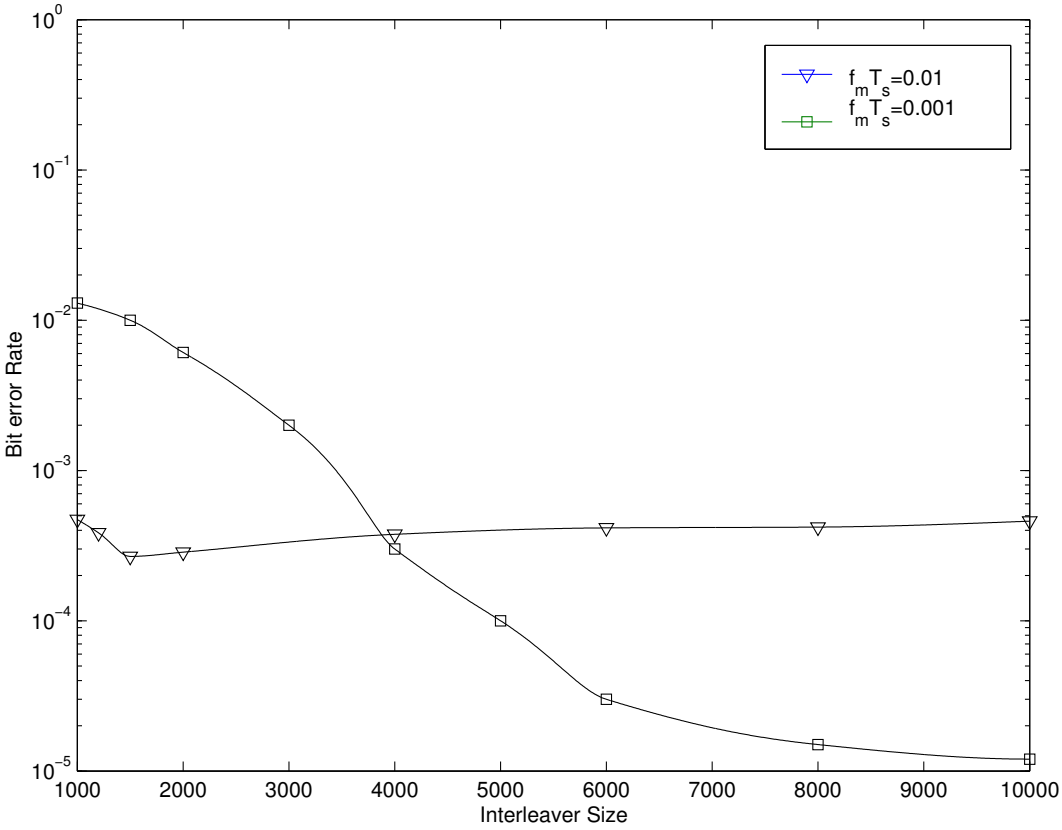


Figure 31: Effect of interleaver size over time-selective fading channels. The SEP-based A-SISO demodulator with the *SNR* stopping criterion is employed. $E_b/N_o = 10\,\text{dB}$.

iterations, the fixed-iteration criterion is the best among them, and the performances of the *SNR* and average-entropy criteria are almost same.

4.4.3 Effects of Interleaver Size

It is a well-known fact that for turbo or "turbo-like" codes a larger interleaving gain can be achieved by using a longer interleaver. Moreover, it is shown in [33] that the time-diversity gain resulting from the use of interleaver on a fading channel grows with an increase in the Doppler spread of channels. Fig. 31 shows the performances of A-SISO receivers with different interleaver sizes, where fading channels with $f_m T_s = 0.001$ and $f_m T_s = 0.01$ are considered. Since convergence oscillations occur over these channels, the SEP-based A-SISO demodulator with the *SNR* iteration-stopping criterion is employed in simulations. When the channel is slow time-variant, e.g., $f_m T_s = 0.001$, the interleaving gain is significant, especially for receivers with large interleaver size. However, on a fast fading channel with

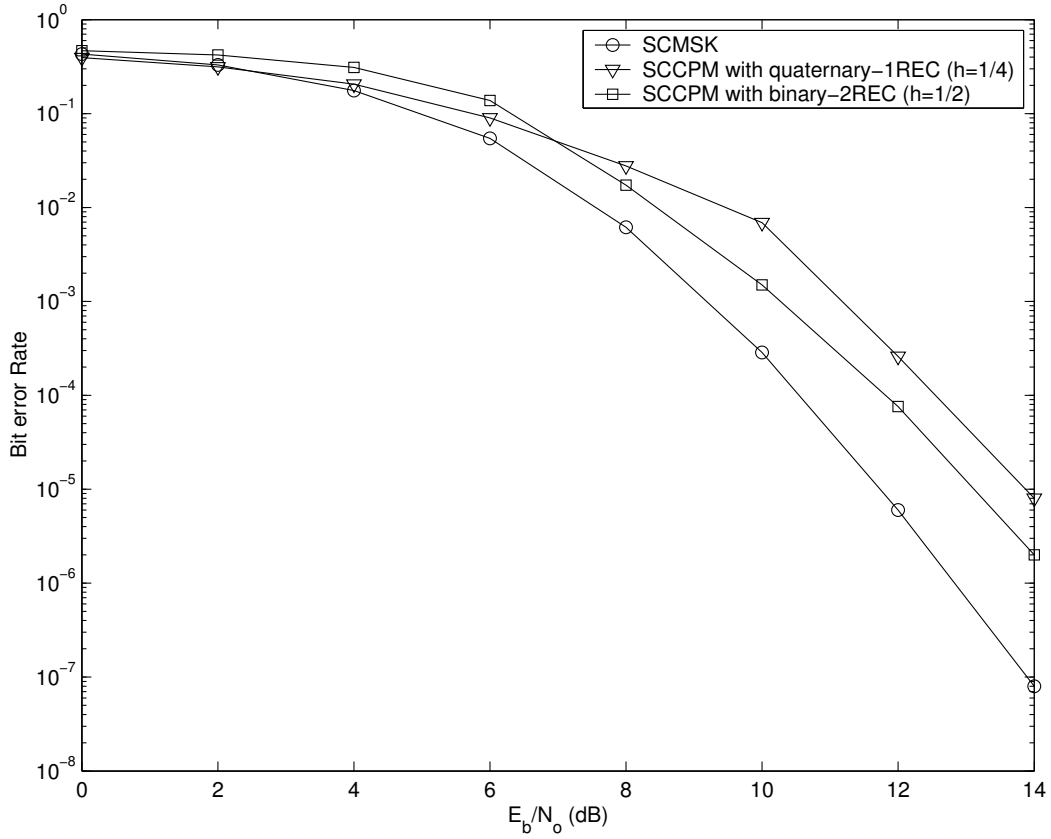


Figure 32: Performances of different SCCPM systems on a fading channel with $f_m T_s = 0.001$. The SEP-based A-SISO demodulator with the *SNR* stopping criterion is employed. $N_x = 4096$.

$f_m T_s = 0.01$, the mean time to cycle slipping is reduced significantly compared to slow fading channels. Therefore, with an increase of interleaver size, the cycle slipping occurs more frequently, and the effect of such phase rotations finally dominates the system performance. As a result, the interleaving gain diminishes gradually as the interleaver size grows. Furthermore, the time-diversity gain of fast fading channels can be effectively ignored.

Although not shown here, it is observed that convergence oscillations appear even when the normalized Doppler frequency is less than 0.001, although a low BER performance still can be achieved by using the SEP-based demodulator with the *SNR* iteration-stopping criterion. Generally speaking, the appearance of convergence fluctuation not only depends on channel dynamics but also on frame length (or interleaver length). For slow fading channels with low phase dynamics, a larger interleaver may be preferred to combat burst errors, which significantly increases the probability of occurrence of cycle slipping within

each frame and further leads to convergence oscillations.

4.4.4 Performance of Different SCCPM Systems

The performance comparison of different SCCPM systems with unknown carrier phase is shown in Fig. 32 and 33, where time-selective fading channels with $f_m T_s = 0.001$ and $f_m T_s = 0.01$ are considered. Besides MSK, two CPM systems with multiple symbol levels or partial response, i.e, quaternary 1REC with $h_{\text{idx}} = 1/4$ and binary 2REC with $h_{\text{idx}} = 1/2$, are included, where LREC denotes a CPM scheme with rectangular frequency pulse having length L. To implement iterative phase synchronization and detection, the SEP-based A-SISO receiver with *SNR* stopping criterion is employed. At $f_m T_s = 0.001$, no error floor is observed at a BER of 10^{-6} for the three systems. When $f_m T_s = 0.01$, carrier phase changes rapidly and, thus, the performance loss owing to imperfect phase tracking is significant. An error floor at a BER of 10^{-4} exists for all considered systems.

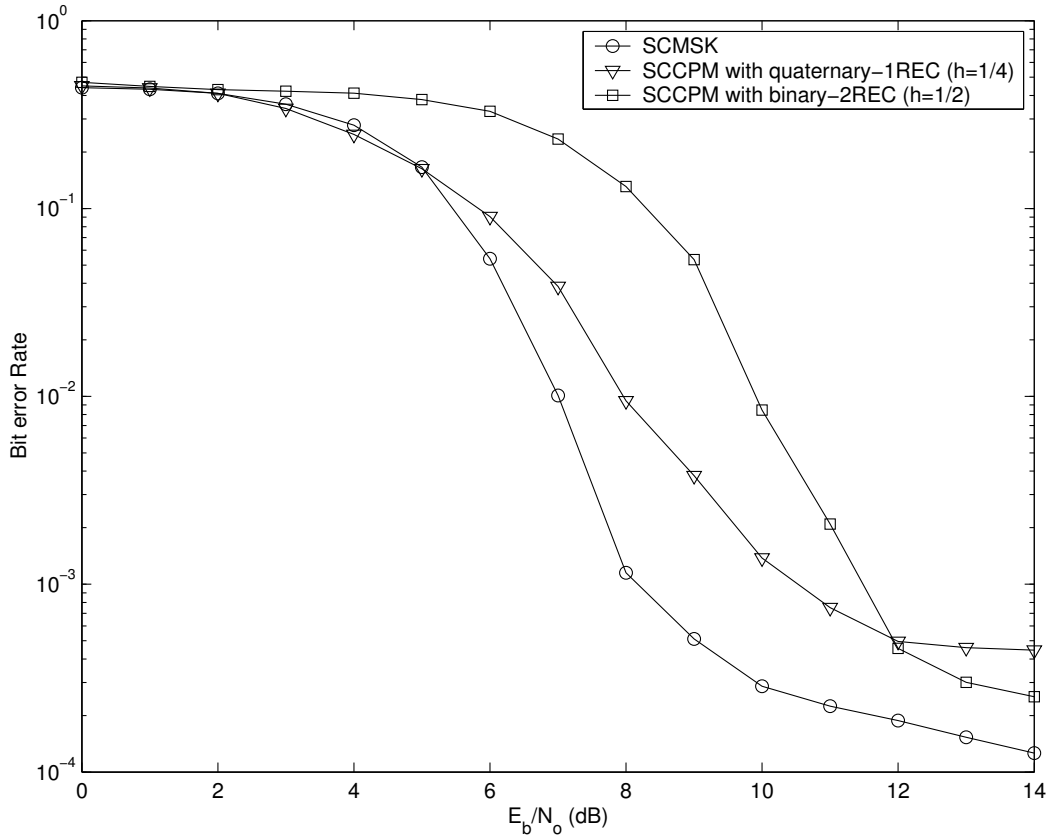


Figure 33: Performances of different SCCPM systems on a fading channel with $f_m T_s = 0.01$. The SEP-based A-SISO demodulator with the *SNR* stopping criterion is employed. $N_x = 2048$.

CHAPTER V

TURBO SYNCHRONIZATION FOR SERIALLY CONCATENATED CPM

In Chapter II, we have proposed a data-aided joint time and phase synchronizer which works well for all CPM schemes and requires less training compared to existing techniques. However, for any CPM data-aided synchronization method, extending it into its decision-directed counterpart is not trivial. Actually, it was reported in [18] and [20] that a decision-directed synchronizer cannot synchronize partial-response CPM signals regardless of the format of applied time and phase estimators because of the existence of false locks. In this chapter, based on the combination of both data-aided and decision-directed techniques, we introduce a mixed method for CPM synchronization, which can effectively prevent the occurrence of false locks. Since we consider serially concatenated CPM and apply the A-SISO algorithms in the iterative receiver, the resulting synchronization is therefore denoted as turbo synchronization. Simulation results show that the proposed turbo synchronizer is robust in acquiring and tracking both time-invariant and time-variant channels at very low signal-to-noise-ratios. The main results in this chapter can be found in [113].

5.1 Turbo Synchronization for SCCPM

5.1.1 System Description

The proposed turbo synchronization scheme is compatible with any CPM FEP and any CPM synchronization method. In this chapter we considered only the FEP based on non-symmetric nonorthogonal exponential expansion (nSnOEE) and the joint timing and phase recovery method presented in Chapter II, because of their superiority in terms of complexity and performance. In the remainder of this chapter, we follows the same notations defined in Chapter II. Recall that based on nSnOEE, the suboptimal FEP requires only N_f filters.

Let $\alpha_l(t)$ be the l-th exponential-shaping filter defined on $[0, T_s]$. With *synchronized* FEP, the sampling rate is $1/T_s$, and the output of l-th FEP filter at the k-th time epoch is

$$y_{l,k} = \exp\{-j\phi'_k\} \times \left\{ y(t) \otimes \alpha_l^*(t)|_{t=kT_s+\tau'_k} \right\}, \tag{104}$$

where τ'_k and ϕ'_k are the sampling instant and phase compensation, respectively. The observation record $\boldsymbol{Y}_k = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_k]$ with $\boldsymbol{y}_k = [y_{1,k}, \cdots, y_{N_f,k}]^T$ consists of the immediate suboptimal sufficient statistics for estimation and detection. With *asynchronous* FEP, an fixed over-sampling rate R_s is typically used and the outputs of the l-th filter are

$$z_{l,m} = y(t) \otimes \alpha_l^*(t)|_{t=mT_s/R_s}. \tag{105}$$

Construct the $N_f \times m$ matrix $\boldsymbol{Z}_m = [\boldsymbol{z}_1, \cdots, \boldsymbol{z}_m]$, where $\boldsymbol{z}_m = [z_{1,m}, \cdots, z_{N_f,m}]^T$. Given τ'_k and ϕ'_k, $\boldsymbol{y}_k$ can be accurately constructed from $\boldsymbol{Z}_m$ using appropriate interpolation techniques [114], [115], [116]. The *asynchronous* FEP is particularly important when PSP-based algorithms are used for joint synchronization and detection, which is elaborated in the next subsection.

For turbo-coded systems, parameter estimation, including channel estimation and synchronization, can be processed either inside or outside of an iterative detection loop. With the inside-loop method, both the parameter estimates and the data decisions are refined through successive iterations by exchanging soft information between the estimators and decoders. With the outside-loop method, estimation is performed only once prior to iterative detection. Typically, the inside-loop method outperforms its outside-loop counterpart in terms of the MSE performance, since refined data decisions can be used for further training. Moreover, the inside-loop method allows to track the time-variant unknown parameters while the latter does not.

In order to ensure robust acquisition and tracking, the proposed turbo synchronization scheme uses the mixture of a preamble, pilots and tentative decisions to drive the time and phase looked loops. The outside-loop data-aided parameter acquisition using the preamble is first executed. Then the inside-loop synchronization is processed by the help of pilot symbols and tentative decisions, using the A-SISO CPM demodulator.

5.1.2 The A-SISO CPM Demodulator

The same iterative receiver as in Fig. 19 is used to implement joint synchronization and SISO CPM demodulation in a PSP fashion. Within the PSP-based algorithms, parameter estimation is performed along with each survivor. Therefore, multiple estimators are normally processed in parallel and multiple tentative estimates are retained at each time epoch. In the CPM tilted-phase trellis, denote the i-th forward survivor at the k-th interval as $\tilde{\mathbf{u}}_1^k(i)$. Note that given a initial start state in trellis encoding, each symbol sequence uniquely corresponds to a continuously connected branch sequence. In the following, for notation conciseness, we denote the survivor $\tilde{\mathbf{u}}_1^k(i)$ as a sequence of survived/selected branches, and $b_k(i)$ as the i-th survived branch at the k-th interval that is associated with one of possible pM^L transmitting CPM signals. The i-th set of forward estimates at the k+1-st time epoch, $\tilde{\Phi}_{k+1}^f(i) = [\tilde{\phi}_{k+1}(i), \tilde{\tau}_{k+1}(i)]^T$, is predicted as

$$\tilde{\Phi}_{k+1}^f(i) = \Upsilon_k\big(\tilde{\Phi}_k^f(i), \tilde{\boldsymbol{y}}_k(i), b_k(i)\big), \quad i = 1, \cdots, N_s, \tag{106}$$

where $\tilde{\boldsymbol{y}}_k(i)$ are the tentative sufficient statistics associated with the i-th forward survived branch $b_k(i)$ and can be constructed from $\boldsymbol{Z}_m$ with *asynchronous* FEP, $\Upsilon_k(\cdot)$ is the parameter prediction function at the k-th interval denoting the equation (25) plus a Kalman filter for computing $\hat{\boldsymbol{\epsilon}}_k$, and N_s is the number of survivors which is unity for the SEP method. Note that if we use *synchronous* FEP to compute $\tilde{\boldsymbol{y}}_k(i)$, the sampling device has to operate in a very high sampling rate, since the difference between multiple tentative sampling instants for different survivors is usually quite small. The resulting *synchronous* FEP is therefore prohibitively expensive.

Denote $e_k(i)$ as the ending state of $b_k(i)$. The entire set of forward branches at the k+1-st time epoch is updated as

$$\big\{b_{k+1}(i)\big\}_{i=1}^{N_s} = \arg \underset{b:e_k(j),\, 1\leqslant j\leqslant N_s}{\text{multiMax}} \Gamma^f(e_k(j))\mathbb{P}\big(\tilde{\boldsymbol{y}}_{k+1}(j), b|\tilde{\Phi}_{k+1}^f(j)\big), \tag{107}$$

where $\{b{:}e_k(j)\}$ denotes the hypothesis branch b starting from the state $e_k(j)$, $\Gamma^f(e_k(j))$ is the forward state metric at the state $e_k(j)$, and $\mathbb{P}\big(\tilde{\boldsymbol{y}}_{k+1}(j), b|\tilde{\Phi}_{k+1}^f(j)\big)$ is the branch metric associated with the branch b. The operator arg multiMax returns the N_s survived branches

with the N_s largest APPs. The forward state metric $\Gamma^f(e_k(j))$ can be computed in a similar way as in the classical BCJR algorithm,

$$\Gamma^f(e_k(j)) = \sum_{b \in \mathcal{B}_k(j)} \Gamma^f(e_{k-1}(l)) \mathbb{P}\big(\tilde{\boldsymbol{y}}_k(l), b \,|\tilde{\Phi}_k^f(l)\big), \tag{108}$$

where the branch set $\mathcal{B}_k(j)$ denotes all possible branches connecting the states $e_{k-1}(l)$ and $e_k(j)$ with $1 \leqslant l \leqslant N_s$. Moreover, the backward survivors and estimates are updated in the same way.

In (107) and (108), when updating the survived branch and state metric, it is assumed that no data information is available and, thus, all possible branches associated with different information symbols must be considered. Recall that within the A-SISO CPM demodulator, the pilot symbols are used to prevent the false locks. In the CPM trellis, in order to determine the transmitting signal in a specified symbol interval, not only must the corresponding modulated information symbol be known, but also the ending trellis state at the previous time stage must be known. Therefore, in the symbol intervals where pilots are transmitted, it is impossible to uniquely determine the transmitting signal, i.e., the associated trellis branch, owing to the uncertainty of the ending (or start) trellis state at the previous (or present) symbol interval. As a result, the multiple branches associated with the pilot have to be considered in the synchronizer. Similar to the PSP approach, the multiple estimators are required to be processed in parallel in the pilot-aided CPM synchronizer. By limiting the number of the estimators and considered branches to N_s, the similar procedure described in (106), (107) and (108) can be used again. The only difference is that the fewer hypothesis branches need to be considered in the survivor selection and metric recursion, since a number of branches that do not correspond to the pilot symbol must be discarded. The resulting computation is, thus, less complex.

5.2 Numerical Results And Discussions

The performance of turbo synchronizer, aided by the preamble and pilot symbols, has been evaluated by simulations for SCCPM systems. The system setup is as follows. At the

transmitter, a pseudorandom interleaver and a rate-1/2 convolutional outer code with generators $[5_8, 7_8]$ are employed. Typical of $\mathcal{M}$-ary response CPM systems, a 4-ary-3RC CPM system is considered with the modulation index $h_{\text{idx}} = 1/2$. Unless otherwise indicated, the receiver uses nSnOEE-based FEP followed by an A-SISO module. For 4-ary-3RC CPM, a three-dimensional nSnOEE is adopted with the frequencies of exponential shaping pulses being $\boldsymbol{f} = 1/T_s \times [0, 0.3, 1.0]$. We also assume for simplicity that frequency offset has been perfectly recovered at receiver.

Note that the PSP-based A-SISO algorithms can be classified into two categories, namely, the forward-only and forward-backward A-SISO algorithms. According to known results in [98], we adopt the forward-only A-SISO algorithm throughout this section because it is more robust than the forward-backward A-SISO algorithms in tracking time-variant parameters.

Finally, owing to the use of the preamble and pilot symbols, the power efficiency of the SCCPM system with turbo synchronization is correspondingly reduced. Denote the ratio between the length of the preamble and the length of data and pilot symbols as R_{pre}. Also let the ratio between the number of the pilots and the length of data symbols be R_{plt}. The total power loss owing to the overhead, i.e., P_{loss} (dB), is given by

$$P_{\text{loss}} = 10 \times \Big(\log_{10}(1 + R_{\text{pre}}) + \log_{10}(1 + R_{\text{plt}}) \Big). \tag{109}$$

5.2.1 BER Performance of Turbo Synchronization

Fig. 34 shows the BER performance of the proposed iterative SCCPM receiver with time-invariant and time-variant timing and phase errors. An AWGN channel is considered. The total number of modulated CPM symbols in a frame is set to 1000, among which the length of preamble is 200 and the number of pilot symbols is 150. Correspondingly, $R_{pre} = R_{plt} = 0.25$ and, thus, the resulting power loss owing to the overhead is approximately 2.0 dB according to (109). For convenience of comparison, the performance of the SCCPM iterative receiver with ideal knowledge of timing and phase information is also included in Fig. 34, where no preamble and pilot symbol are used.

The two typical operating scenarios (OS) are considered for evaluating the performance

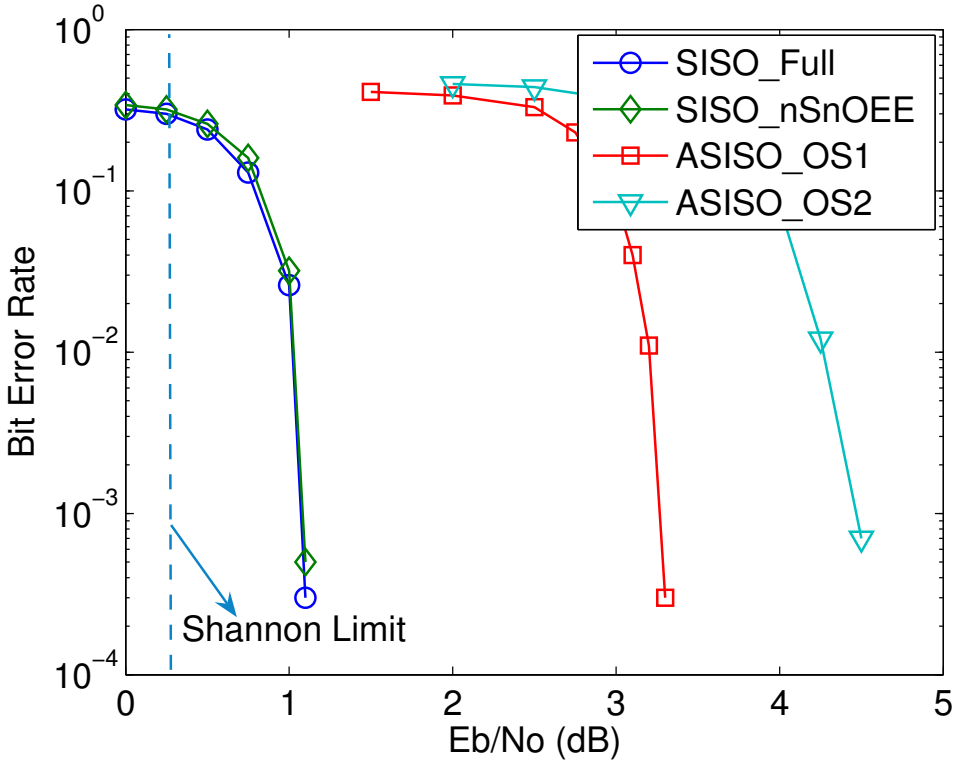


Figure 34: BER performance of the proposed turbo synchronizer.

of turbo synchronization. In the first scenario, the unknown parameters are assumed constant over a frame. In the second scenario, the time-variant parameter offsets are assumed and are modeled as discrete-time Wiener processes. The rms of time and phase increments in each symbol interval are set as $\sigma_\tau = 0.001T_s$ and $\sigma_\phi = 2^\circ$, respectively. Note that $\sigma_\tau > 0.001T_s$ corresponds to severe operating conditions and is rarely observed in practice. In both scenarios, the initial errors are set to be $\tau_0 = 0.4T_s$ and $\phi_0 = 0.6\pi$ in order to examine the acquisition performance in the worst-case. Fig. 34 shows that when the time and phase offsets are perfectly known at the receiver, the performance loss owing to the use of nSnOEE-based suboptimal FEP is negligible compared to the receiver with full-complexity FEP. The performance of the ideal receivers is approximately 1.0 dB away from the Shannon limit [117]. When acquiring and tracking unknown time and phase information, the performance loss owing to turbo synchronization is less than 2.4 dB in the first scenario, and less than 3.5 dB in the second scenario. Note that in both scenarios, 2.0 dB out of the total loss is due to the use of preamble and pilot symbols.

In Chapter IV, for joint phase synchronization and CPM demodulation, an oscillatory BER convergence behavior was observed on iterative receivers using the A-SISO algorithms, because of the existence of cycle slips in phase-locked loops. However, in the proposed timing and phase synchronizer, no such convergence fluctuation is found owing to the use of preamble and pilot symbols, which effectively remove the cycle slips in both time/phase locked loops.

5.2.2 Effects of Incorrect Modeling

In Fig. 34, perfect knowledge of the channel statistics, including the variance of channel noise, the probability distribution and covariance of time and phase jitters, is assumed within the turbo synchronizer. However, such information is known only approximately in practice. It was shown in Chapter II that the nSnOEE-based noniterative data-aided synchronizer is stable with an incorrect channel model. In particular, the variance of estimate errors converges to zero on time-invariant channels, but at a slower convergence rate when compared to the case of perfect channel modeling. In Fig. 35 we study the effect of channel modeling errors on the proposed turbo synchronizer. The following scenario is considered: 1) a time-invariant channel with the initial time and phase errors being $\tau = 0.4T_s$ and $\phi = 0.6\pi$; 2) the estimate of noise power being $\hat{N}_o = E_b$ no matter what the true value of N_o is, i.e., E_b/N_o is always assumed to be 0 dB. Moreover, the frame length and the configuration of preamble and pilots are same as in Fig. 34. It is worth noting that the estimate of noise variance is required for both metric computation in MAPSD and Kalman filtering in synchronization. In order to isolate the effect of incorrect noise variance estimation on synchronization from that on MAPSD, we simply assume the noise variance estimate used in MAPSD is perfect. It was reported in [102] that the performance loss owing to the SNR mismatch in MAPSD was negligible when the SNR offset fell in $[-3\,\text{dB}, 6\,\text{dB}]$.

Fig. 35 shows that, for time-invariant channels, the effect of incorrect channel modeling is negligible for the turbo synchronizer, while it is more significant for the noniterative receiver. The reason is as follows. For a noniterative receiver, the equivalent length of training sequence, including preamble, pilots and decisions, is limited by the frame length, and so

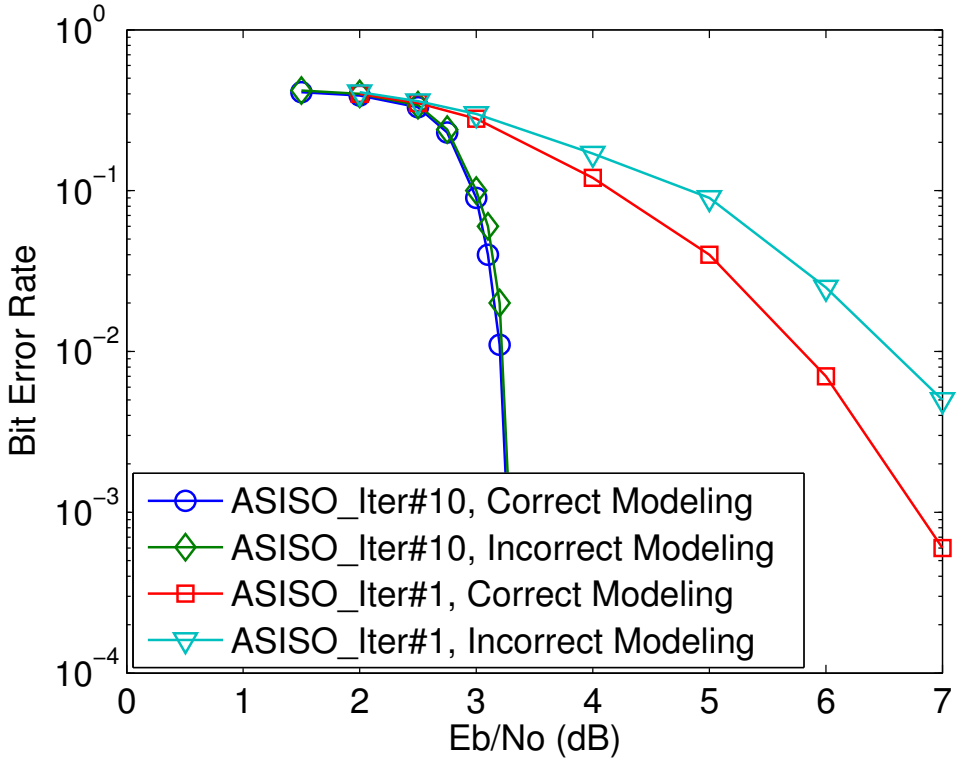


Figure 35: The effects of incorrect channel modeling.

is the synchronization performance. Along with the increase of E_b/N_o from 0 dB to 7 dB, the ratio of the assumed noise variance to the actual noise variance rises correspondingly and, thus, the convergence rate of parameter estimation decreases significantly and so does the BER performance. For the turbo receiver, synchronization is iteratively processed in a decision-directed manner and, thus, the equivalent length of training sequence increases proportionally through the iterations. As a result, the performance of the turbo synchronizer improves though at a reduced convergence rate. Once the equivalent training sequence is sufficiently long or the number of iterations is sufficiently large, the estimation errors are small enough and have a negligible effect on MAPSD.

CHAPTER VI

NONCOHERENT SEQUENCE DETECTION ON RAYLEIGH FADING CHANNELS

In this chapter we compares noncoherent and coherent detection on Rayleigh flat fading channels. Based on the first-order Gauss-Markov channel model, we derive the error probabilities of noncoherent detection for full-response CPM and DPSK, on both time-invariant and time-variant channels. This result explicitly shows how the performance of noncoherent detection is affected by detector complexity, fading rate and operating SNR. Furthermore, using this analysis, we propose a hybrid system which switches between a noncoherent detector and a coherent detector with imperfect channel estimates, in order to optimize the overall system performance, in terms of service quality, receiver complexity and power efficiency. The main results in this chapter can be found in [118].

6.1 System Model and Problem Formulation

The received complex envelope on a flat Rayleigh fading channel is

$$y_k = h_k x_k + n_k, \qquad k = 1, 2, \cdots, N_x, \tag{110}$$

where x_k is the transmitted symbol having energy E_s and duration T_s, h_k is the complex channel gain, n_k represents a zero-mean complex white Gaussian noise sample with variance N_o, and N_x is the frame length. We assume that the complex channel gain h_k is constant over one symbol interval and has average power equal to $\mathbb{E}\{h_k h_k^*\} = \Omega_h$. If the channel actually varies over the symbol interval T_s, this assumption leads to suboptimal sufficient statistics. Moreover, we assume that the transmitted signal x_k, the channel h_k and the noise n_k are jointly independent. Finally, we denote the modulation alphabet and its size as $\mathcal{X}$ and $|\mathcal{X}|$, respectively.

Table 4: The practical values of the coefficient ρ on time-variant fading channels.

	$f_m = 0.03$	$f_m = 0.01$	$f_m = 0.003$	$f_m = 0.001$
$M_r = 2$	$\rho \approx 0.9911$	$\rho \approx 0.9990$	$\rho \approx 0.9999$	$\rho > 0.9999$
$M_r = 3$	$\rho \approx 0.9841$	$\rho \approx 0.9982$	$\rho \approx 0.9998$	$\rho > 0.9999$
$M_r = 4$	$\rho \approx 0.9772$	$\rho \approx 0.9975$	$\rho \approx 0.9998$	$\rho \approx 0.9999$
$M_r = 5$	$\rho \approx 0.9704$	$\rho \approx 0.9967$	$\rho \approx 0.9997$	$\rho \approx 0.9999$

6.1.1 Channel Model

The channel fading dynamics are modelled as a first-order Gauss-Markov (GM) process[1],

$$h_{k+1} = \rho h_k + v_k, \tag{111}$$

where the coefficient $\rho \in [0, 1]$ characterizes the fading rate, and v_k is a zero-mean white complex Gaussian noise sample with variance $\Omega_h(1-\rho^2)$. Note that $\{h_1, h_2, \cdots, h_k\}$ are independent of v_k, and $\{v_1, \cdots, v_{N_x}, n_1, \cdots, n_{N_x}\}$ are jointly independent. Obviously, the time variation of h_k increases with increasing ρ. However, the practical range of ρ is still an open question.

For a 2-D isotropic scattering environment, the time-variant fading channel gain h_k has the *normalized* autocorrelation function [119]

$$R_{h,\text{ideal}}(l) = \frac{\mathbb{E}\{h_{k+l}h_k^*\}}{\mathbb{E}\{h_k h_k^*\}} = J_0(2\pi l f_m), \tag{112}$$

where f_m is the normalized maximum Doppler frequency and $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind. However, with the simplified first-order GM model given in (111), the *normalized* autocorrelation of h_k is

$$R_h(l) = \frac{\mathbb{E}\{h_{k+l}h_k^*\}}{\mathbb{E}\{h_k h_k^*\}} = \frac{\mathbb{E}\{\rho^l h_k h_k^*\}}{\mathbb{E}\{h_k h_k^*\}} = \rho^l. \tag{113}$$

To determine the appropriate value of ρ for each Doppler frequency, a straightforward

[1]The first-order Gauss-Markov channel model has been widely used in the literature when analyzing BER, cutoff rate or capacity of temporally correlated fading channels [66, 67, 70].

criterion is adopted,

$$\begin{aligned}\rho(f_m) &= \arg\min_{\rho\in[0,1]} \sum_{l=-M_r}^{+M_r} |R_{h,\text{ideal}}(l) - R_h(l)|^2, \\ &= \arg\min_{\rho\in[0,1]} \sum_{l=-M_r}^{+M_r} |J_0(2\pi l f_m) - \rho^l|^2, \end{aligned} \tag{114}$$

where M_r determines the width of the sliding observation window. For a given f_m, the coefficient ρ varies with M_r. Table I shows the optimal values of ρ for different normalized Doppler frequencies $f_m T_s$ on the order of 10^{-2} and 10^{-3}, which are typically considered to be rapidly time-invariant fading channels in most commercial and military applications. More important, even when the channel is rapidly time-varying, the coefficient ρ is still close to 1, typically lying in the interval $[0.999, 0.99999]$. This fact plays an important role in the performance analysis of coherent and noncoherent sequence detection using the first-order GM channel model, especially for practical values of E_s/N_o lying in the interval $[10\,\text{dB}, 30\,\text{dB}]$. More details are given in Subsection 6.1.3 and Section 6.2. Also note that in the remainder of this chapter, we assume $\Omega_h = 1$ for simplicity of notation and derivation.

6.1.2 Coherent and Noncoherent Sequence Detection

In coherent detection, the channel gain must be estimated at the receiver. When the channel gain is known perfectly at the receiver, the criterion for maximum likelihood coherent sequence detection is [117]

$$\hat{\mathbf{s}} = \arg\min_{\tilde{x}\in\mathcal{X}} \sum_{k=1}^{N_x} \|y_k - h_k\tilde{x}_k\|^2. \tag{115}$$

where $\mathbf{y} = [y_1, \cdots, y_{N_x}]^T$ spans the entire received signal record, and $\hat{\mathbf{x}}$ and $\tilde{\mathbf{x}}$ represent the estimate and hypothesis of the transmitted symbol sequence, respectively. In (115), $\|A\|^2$ is equal to the inner product $\langle A, A\rangle$; for discrete-time samples $\langle A, B\rangle = AB^*$, while for continuous-time signals $\langle A, B\rangle = \frac{1}{T_s}\int_{T_s} A(t)B^*(t)dt$.

When the channel gain is approximately known at receiver, the criterion in (115) is no longer optimal in the sense of maximizing the likelihood function, since the channel estimation error is colored. However, in practice the criterion in (115) is still widely used for

coherent sequence detection with imperfect channel knowledge. The effect of this mistreatment on the overall detection performance is discussed in the next subsection.

Noncoherent detection does not require an estimate of the channel gain at receiver. Assuming perfect knowledge of the probability density function of the channel gain at the receiver, the criterion for noncoherent sequence detection is

$$\hat{\mathbf{x}} = \arg\max_{\tilde{x}\in\mathcal{X}} \int_{\mathcal{H}^{N_x}} \exp\left\{-\frac{1}{N_o}\sum_{k=1}^{N_x} \|y_k - h_k\tilde{x}_k\|^2\right\} p(\mathbf{h})\, d\,\mathbf{h}, \tag{116}$$

where $\mathcal{H}^{N_x}$ is an N_x-dimensional parameter space for the channel gain $\mathbf{h} = [h_1, \cdots, h_{N_x}]^T$. For a time-variant fading channel, the evaluation of the argument in (116) is extremely complex, if not impossible, because an N_x-dimensional integral must be computed. When the channel is time-invariant, i.e., $h = h_1 = \cdots = h_{N_x}$, the N_x-dimensional integral reduces to a single integral that can be evaluated as [75]

$$\begin{aligned}\int_{\mathcal{H}} \exp\left\{-\frac{1}{N_o}\sum_{k=1}^{N_x} \|y_k - h\tilde{x}_k\|^2\right\} p(h)dh &\propto \int_{\mathcal{H}} \exp\left\{\frac{2h}{N_o}\sum_{k=1}^{N_x} \mathrm{Re}\{\langle y_k, \tilde{x}_k\rangle\}\right\} p(h)\, dh, \\ &\propto I_0\left(\sqrt{\frac{2\left\|\sum_{k=1}^{N_x}\langle y_k, \tilde{x}_k\rangle\right\|}{N_o}}\right), \end{aligned} \tag{117}$$

where we further assume the phase information of h is uniformly distributed on $[0, 2\pi]$ and $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind. Since $I_0(\cdot)$ is a monotonically increasing function, the criterion in (117) simplifies to [75]

$$\hat{\mathbf{s}} = \arg\max_{\tilde{x}\in\mathcal{X}} \left\|\sum_{k=1}^{N_x}\langle y_k, \tilde{x}_k\rangle\right\|. \tag{118}$$

We assume that all valid transmitted sequences appear with equal probability. For both coherent and noncoherent sequence detection, the symbol or bit error probability is thus upper bounded by [117]

$$P_e \leqslant \sum_{\hat{\mathbf{x}}\neq\mathbf{x}} \mathcal{W}_{\hat{\mathbf{x}}}\mathbb{P}(\hat{\mathbf{x}}|\mathbf{x}), \tag{119}$$

where $\mathbb{P}(\hat{\mathbf{x}}|\mathbf{x})$ denotes the probability that $\hat{\mathbf{x}}$ is selected according to a decision criterion while $\mathbf{x}$ is actually transmitted, and $\mathcal{W}_{\hat{\mathbf{x}}}$ refers to the number of bit or symbol errors in $\hat{\mathbf{x}}$. Typically, P_e is dominated by $\mathbb{P}(\hat{\mathbf{x}}_{free}|\mathbf{x})$ with $\hat{\mathbf{x}}_{free}$ being the error sequences at the

minimum (or *free*) Euclidean distance from the true symbol sequence, i.e.,

$$P_e \approx K_e \mathbb{P}(\hat{\mathbf{x}}_{free}|\mathbf{x}), \tag{120}$$

where K_e is a constant determined by the trellis structure. Therefore, in the remainder of the chapter, we focus on the evaluation of $\mathbb{P}(\hat{\mathbf{x}}_{free}|\mathbf{x})$ and denote it as P_d. Moreover, without loss of generality, we assume that within any error sequence the first error occurs in the first symbol interval. If the corresponding diverged trellis path merges with the true path after m symbol intervals, the length-m error event is denoted as $\hat{\mathbf{x}} = [f_1, \cdots, f_m, x_{m+1}, \cdots, x_{N_x}]^T$, where f_i is the symbol associated with the i-th epoch. Note that f_i might be equal to x_i except at $i = 1$. For $\hat{\mathbf{x}}_{free}$ corresponding to a length-m error event, the pairwise error probability P_d under coherent sequence detection is

$$P_d = \mathbb{P}\left(\sum_{k=1}^{m} \|y_k - \hat{h}_k f_k\|^2 \leq \sum_{k=1}^{m} \|y_k - \hat{h}_k x_k\|^2\right), \tag{121}$$

while the error probability P_d under noncoherent sequence detection is

$$P_d = \mathbb{P}\left(\left\|\sum_{k=1}^{N_x} \langle y_k, x_k\rangle\right\|^2 \leq \left\|\sum_{k=1}^{m} \langle y_k, f_k\rangle + \sum_{k=m+1}^{N_x} \langle y_k, x_k\rangle\right\|^2\right). \tag{122}$$

Let $\mathcal{M}_{N_x}^{coh}$ and $\mathcal{M}_{N_x}^{ncoh}$ denote the arguments in equations (115) and (118), respectively. For coherent detection, $\mathcal{M}_{N_x}^{coh}$ can be recursively computed as $\mathcal{M}_{N_x}^{coh} = \mathcal{M}_{N_x-1}^{coh} + \|y_{N_x} - h_{N_x}\tilde{x}_{N_x}\|^2$. Therefore, by employing proper tree-pruning rules as in the Viterbi algorithm, maximum likelihood sequence detection can be processed with a complexity that increases linearly with the frame length, i.e., a complexity on the order $\mathcal{O}(N_x|\mathcal{X}|)$. In contrast, the computation of the metric $\mathcal{M}_{N_x}^{ncoh}$ in noncoherent sequence detection cannot be processed in a recursive fashion. Consequently, an exhaustive search must be performed to find the optimal sequence and, thus, the computation complexity increases exponentially with the frame length, i.e., a complexity on the order $\mathcal{O}(|\mathcal{X}|^{N_x})$. If we consider a width-N sliding observation block with $m < N \ll N_x$ and perform the block-by-block noncoherent detection, the complexity of the resulting noncoherent detector is now on the order $\mathcal{O}(\frac{N_x}{N}|\mathcal{X}|^N)$, which is normally acceptable in practice by properly selecting the observation width N.

Correspondingly, the pairwise error probability P_d in (122) must be revised as

$$P_d = \mathbb{P}\left(\left\|\sum_{k=1}^{N}\langle y_k, x_k\rangle\right\|^2 \leq \left\|\sum_{k=1}^{m}\langle y_k, f_k\rangle + \sum_{k=m+1}^{N}\langle y_k, x_k\rangle\right\|^2\right). \tag{123}$$

6.1.3 Error Probability of Coherent Sequence Detection

For coherent sequence detection with either perfect or imperfect channel knowledge, the pairwise error probability is well known and can be computed approximately by using various approaches. In this subsection, we summarize existing main results and use them as a baseline to compare with the performance of noncoherent sequence detection on both time-invariant and time-variant fading channels. In the following review, we limit our attention to modulation schemes having length-2 error events, including full-response CPM and DPSK.

When the channel is known perfectly at receiver, the pairwise error probability of the length-2 *free* distance error event on time-invariant channels ($\rho = 1$) is [120]

$$P_d \approx \frac{N_o}{\|D_1\|^2 + \|D_2\|^2}, \tag{124}$$

and for time-variant channels ($0 \leq \rho < 1$)

$$P_d \approx \frac{3N_o^2}{\|D_1\|^2\|D_2\|^2(1-\rho^2)}, \tag{125}$$

where $D_1 = x_1 - f_1$ and $D_2 = x_2 - f_2$. When $\rho = 0$, the full time-diversity of order 2 can be achieved; otherwise, only a partial diversity gain is available on time-correlated channels. For practical values of ρ, i.e., $\rho \in [0.999, 1)$, the approximation in (125) is poor and the pairwise error probability is more accurately computed as

$$P_d \approx \frac{N_o(1 - K_\rho(1-\rho))}{\|D_1\| + \|D_2\|^2}, \tag{126}$$

where K_ρ is a positive constant determined by the signal constellation, and $K_\rho(1-\rho)$ is typically very small and can be simply ignored in computing P_d.

When the channel gain is known approximately at the receiver, the pairwise error probability P_d on time-invariant channels is approximately [61]

$$P_d \approx \frac{N_o + (1+K_h)P_h E_s}{\|D_1\|^2 + \|D_2\|^2}, \tag{127}$$

where P_h is the variance of channel estimation error, $N_o+P_hE_s$ is the spectral density of the equivalent noise $(h-\hat{h})x_k+n_k$, and the additional term $K_hP_hE_s$ is due to the suboptimality of the decision criterion in (115) made by incorrectly assuming that the equivalent noise is white. It is shown in Appendix D that $0 \leq K_h \leq 1$, depending on the trellis structure and signal constellation. For time-variant channels with ρ assuming values of practical interest, the pairwise error probability P_d is the same as (127). In this case, P_d is related to ρ only through the fact that the performance of the channel estimator, i.e., P_h, is influenced by the fading rate. On time-invariant channels, P_h can be made arbitrarily small given sufficient training; however, on time-variant channels, the performance of linear channel estimation assuming an infinitely long training sequence is lower bounded by

$$P_h \geqslant \frac{(E_s-N_o)(1-\rho^2)+\left((E_s-N_o)^2(1-\rho^2)^2+4E_sN_o(1-\rho^2)\right)^{1/2}}{2E_s}, \tag{128}$$

as shown in Appendix E.

In previous work, the performance of coherent detection was evaluated based on pilot-assisted channel estimation, considering both optimal pilot placement and power allocation [63], [66]-[68]. In this work, we focus on joint estimation and detection (JED) which yields a higher power efficiency or data rate by eliminating training overhead. Note that using JED makes a fair comparison between coherent and noncoherent detection, since both are processed without training. Optimal maximum likelihood sequence estimation and detection can be implemented in a two-step procedure, i.e., first estimating unknown channel gains for each possible data sequence and then selecting the data sequence and its associated parameter estimates having maximum conditional likelihood as the final decision. Unfortunately, the complexity of the resulting scheme is prohibitive. To reduce complexity, decision-directed (DD) techniques are often used in practice, where the estimation is limited to a small set of well-selected sequences (or tentative decisions) [60], [37]. Normally, the estimation error variance using the DD method is greater than P_h in (128) due to possible error propagation within decision feedback mechanism [124], [125]. Moreover, the evaluation of P_h in (128) assumes ideal channel statistics at the linear MMSE channel estimator,

including the initial error variance, noise power spectral density and the fading rate. However, in most practical applications these parameters are known only approximately, leading to further performance degradation in the channel estimates [95], [126]. An exact analytical evaluation of the effect of these nonidealities is difficult, if not impossible. In the remainder of the paper, we set $K_h = 1$ for simplicity and assume that it includes all effects caused by suboptimal channel estimation, although these effects are still underestimated as compared to the simulation results in [95], [124]-[126]. This assumption normally yields a best-case performance of coherent detection with practical channel estimation, and makes it possible to conduct an analytical comparison between noncoherent and coherent detection.

6.2 Noncoherent Sequence Detection

Noncoherent sequence detection does not require explicit channel estimation, although its computational complexity is relatively high. It is well known that, on time-invariant channels, the performance of noncoherent CPM detection can closely approach that of coherent CPM detection provided that the sliding observation width, N, is long enough. However, the effect of channel time-variation and a limited-width observation block on the performance of noncoherent detection is an open problem, as is the design of modulation schemes to optimize detection performance. The following analysis addresses answer these questions.

6.2.1 Distribution of Hermitian Quadratic Form

Computing the error probability in (123) is not straightforward. To facilitate the computation, we introduce two propositions, which were first derived by Turin to compute the probability distribution of a random variable of Hermitian quadratic form [121]. The proofs for the propositions are omitted here. The readers of interests can find them and some interesting applications in [122], [123].

Proposition 1: For any random variable r with probability density function $p(x)$ and characteristic function $\psi_r(z)$, the probability $\mathbb{P}(r < 0)$ can be computed as

$$\mathbb{P}(r < 0) = -\frac{1}{2\pi j} \int_{-\epsilon - j\infty}^{-\epsilon + j\infty} \frac{\psi_r(z)}{z} dz, \tag{129}$$

where ϵ is an infinitely small positive number.

Proposition 2: If a random variable r has the Hermitian quadratic form $x = \mathbf{a}^H\mathbf{A}\mathbf{a}$, where $\mathbf{a}$ is an N_a-dimensional complex Gaussian random vector with zero mean and covariance $\mathbf{R}_a$, and $\mathbf{A}$ is an arbitrary $N_a \times N_a$ Hermitian matrix, then the characteristic function of r is

$$\psi_r(z) = \frac{1}{\det(\mathbf{I} - z\mathbf{A}\mathbf{R}_a)} = \prod_{i=1}^{N_a} \frac{1}{1 - z\lambda_i}, \tag{130}$$

where $\lambda_1, \cdots, \lambda_{N_a}$ are the eigenvalues of $\mathbf{A}\mathbf{R}_a$. Moreover, using *Proposition* 1, the probability $\mathbb{P}(r < 0)$ is

$$\begin{aligned} \mathbb{P}(r < 0) &= -\frac{1}{2\pi j} \int_{-\epsilon - j\infty}^{-\epsilon + j\infty} \frac{1}{z} \prod_{i=1}^{N_a} \frac{1}{1 - z\lambda_i} dz, \\ &= \sum_{\lambda_k < 0} \text{Res}\left(-\frac{1}{z} \prod_{i=1}^{N_a} \frac{1}{1 - z\lambda_i}, \lambda_k \right), \end{aligned} \tag{131}$$

where $\text{Res}(f(z), \alpha)$ denotes the residue of the function $f(z)$ at the point α.

6.2.2 On Time-invariant Channels

Based on (123), the pairwise error probability for a length-2 error-event can be further expressed as

$$P_d = \mathbb{P}\left(\left\| \sum_{k=1}^{N} (hE_s + n'_k) \right\|^2 - \left\| \sum_{k=1}^{2} (\alpha_k hE_s + n''_k) + \sum_{k=3}^{N} (hE_s + n'_k) \right\|^2 \leq 0 \right), \tag{132}$$

where $\alpha_k = \langle x_k, f_k \rangle / E_s$ is the correlation between the signal x_k and f_k and, thus, $D_k = 2E_s(1 - \text{Re}\{\alpha_k\})$; n'_k and n''_k are defined as $n'_k = n_k x_k^*$ and $n''_k = n_k f_k^*$, respectively, and the variances of n'_k and n''_k are $N_o E_s$. Now define $\mathbf{a} = \left[\sum_{k=1}^{2} hE_s + n'_k, \sum_{k=1}^{2} \alpha_k hE_s + n''_k, \sum_{k=3}^{N} hE_s + n'_k \right]^T$ with covariance

$$\mathbf{R}_a = \begin{pmatrix} 4E_s^2 + 2N_o E_s & (\alpha_1^* + \alpha_2^*)(2E_s^2 + N_o E_s) & 2(N-2)E_s^2 \\ (\alpha_1 + \alpha_2)(2E_s^2 + N_o E_s) & \|\alpha_1 + \alpha_2\|^2 E_s^2 + 2N_o E_s & (\alpha_1 + \alpha_2)(N-2)E_s^2 \\ 2(N-2)E_s^2 & (\alpha_1^* + \alpha_2^*)(N-2)E_s^2 & (N-2)^2 E_s^2 + (N-2)N_o E_s \end{pmatrix} \tag{133}$$

The decision variable in (132) is defined as $r = \mathbf{a}^H(\mathbf{Q}_1 - \mathbf{Q}_2)\mathbf{a}$, where

$$\mathbf{Q}_1 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \qquad \mathbf{Q}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}. \tag{134}$$

and, correspondingly, $P_d = \mathbb{P}(r \leq 0)$. The eigenvalues of $(\mathbf{Q}_1 - \mathbf{Q}_2)\mathbf{R}_y$ can be computed as

$$\frac{\lambda_{1,2}}{2E_s} = BE_s \pm \left(B^2E_s^2 + BNE_sN_o\right)^{1/2}, \quad \text{and} \quad \lambda_3 = 0, \tag{135}$$

where $B = (N-2)(1 - \frac{1}{2}\mathrm{Re}\{\alpha_1 + \alpha_2\}) + 1 - \frac{1}{4}\|\alpha_1 + \alpha_2\|^2$. At high E_s/N_o, $BE_s \gg NN_o$ and applying the approximation $(t_1^2 + t_1t_2)^{\frac{1}{2}} \approx t_1 + \frac{1}{2}t_2$ given $t_1 \gg t_2 > 0$, the pairwise error probability P_d can be approximated as

$$\begin{aligned} P_d &\approx \frac{NN_o}{(N-2)(4 - 2\mathrm{Re}\{\alpha_1 + \alpha_2\})E_s + (4 - \|\alpha_1 + \alpha_2\|^2)E_s} \\ &= \frac{NN_o}{(N-2)(\|D_1\|^2 + \|D_2\|^2) + (4 - \|\alpha_1 + \alpha_2\|^2)E_s}. \end{aligned} \tag{136}$$

Apparently, the performance gap between noncoherent and coherent sequence detection diminishes with increasing N. Moreover, given the *free* distance of $\|D_1\|^2 + \|D_2\|^2$, the error probability is minimized at $\mathrm{Im}\{\alpha_1 + \alpha_2\} = 0$. For both differential binary PSK (DBPSK) and minimum shift-keying (MSK) modulation, $\|D_1\|^2 + \|D_2\|^2 = 4E_s$ and $\alpha_1 = -\alpha_2$. Thus, their error probabilities are same and equal to

$$P_d^{\mathrm{MSK}} \approx P_d^{\mathrm{DBPSK}} \approx \frac{NN_o}{4(N-1)E_s}, \tag{137}$$

which is a minimum among all trellis-based schemes with a length-2 error event and a *free* distance of $4E_s$. Fig. 36 compares the pairwise error probabilities of MSK with noncoherent detection and coherent detection having imperfect channel estimation. We consider both coarse and fine channel estimation with $P_h = 0.01$ and $P_h = 0.0001$, respectively. With coarse channel estimation, the detection performance is dominated by the poor channel estimation performance and exhibits an error floor. However, the noncoherent detector does not exhibit a noticeable error floor even at high E_s/N_o and, with increase in the observation width, N, its error rate performance closely approaches that of coherent detection with fine channel estimation.

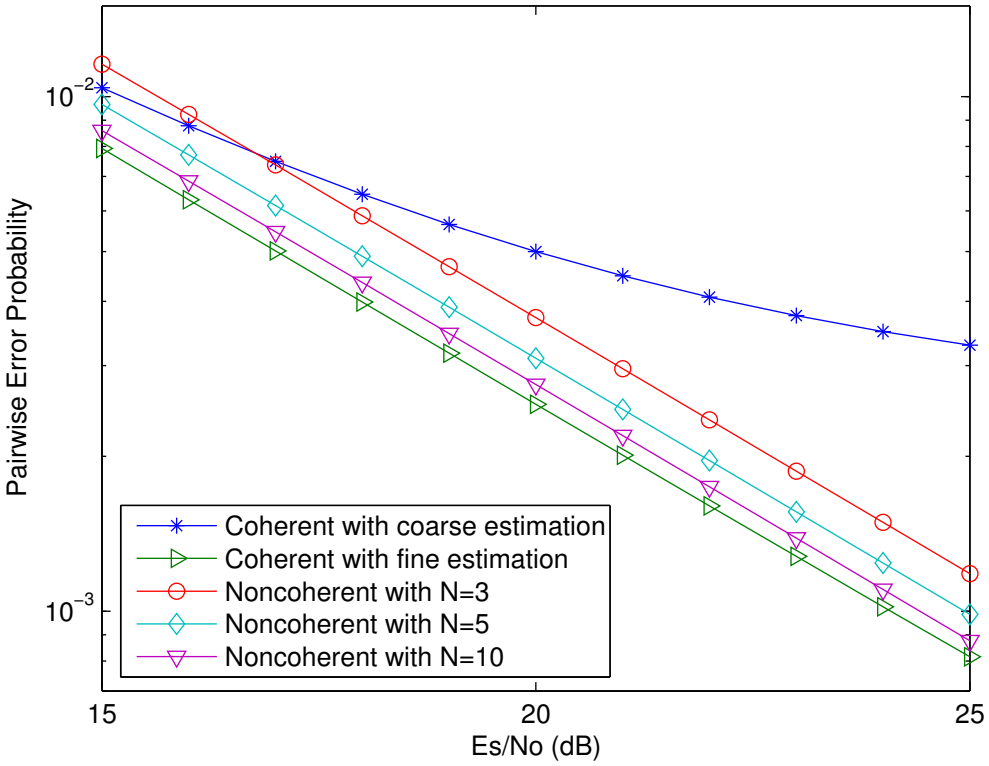


Figure 36: Performance of MSK noncoherent and coherent detection on time-invariant fading channels.

The expression for P_d in (136) is also valid for any full-response CPM scheme. Recall that for coherent detection of full-response CPM with a specific frequency shaping pulse, the optimal modulation index h_{idx}, in terms of maximizing *free* distance, is fixed provided that the observation length is greater than 2 [4]. However, for noncoherent detection, it was shown in [81] via simulations that the optimal modulation index varies with different observation block widths. Based on (136), we confirm this observation in Fig. 37, where the normalized equivalent *free* distance is defined as $\frac{N-2}{N}(4-2\text{Re}\{\alpha_1+\alpha_2\})+\frac{4-\|\alpha_1+\alpha_2\|^2}{N}$ for noncoherent detection and $4-2\text{Re}\{\alpha_1+\alpha_2\}$ for coherent detection, and a rectangular frequency shaping pulse is considered. For example, the optimal modulation indices with noncoherent detection are $h_{\text{idx}}=0.763$ at $N=3$ and $h_{\text{idx}}=0.733$ at $N=5$, yielding 2.03 dB and 1.11 dB power loss, respectively, compared to coherent detection with perfect channel information and the optimal modulation index $h_{\text{idx}}=0.715$.

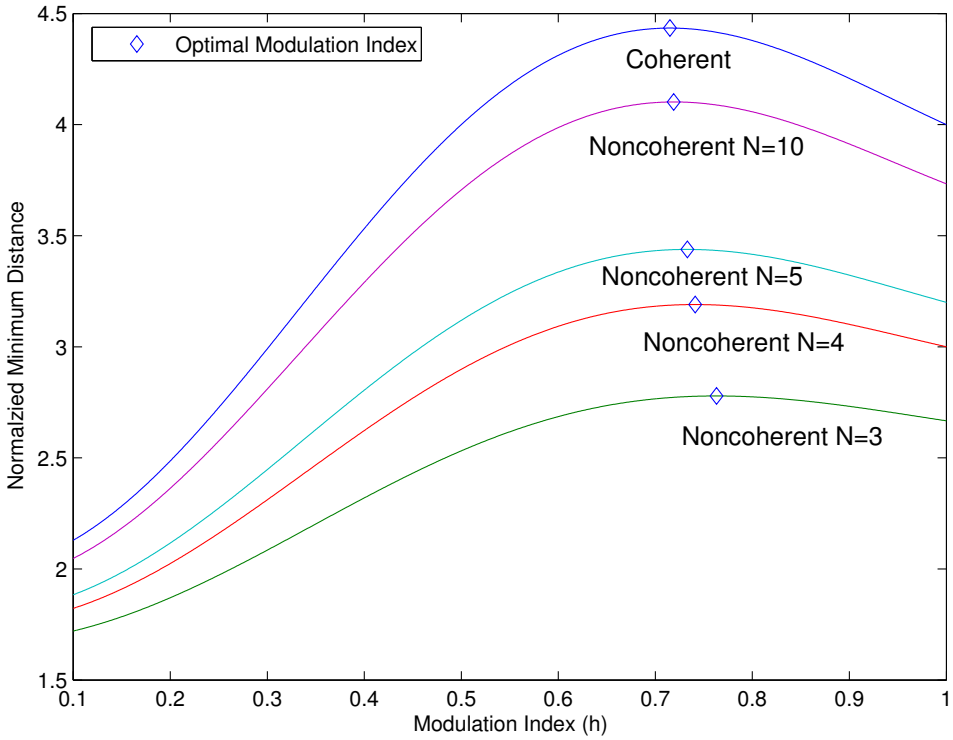


Figure 37: Optimal modulation indices of full-response CPM with rectangular frequency shaping pulse for noncoherent and coherent detection on time-invariant fading channels.

6.2.3 On Time-variant Channels

Strictly speaking, the criterion for noncoherent sequence detection in (118) is optimal only if the channel is time-invariant. However, no practical fading channel is absolutely constant even over a short observation block. In such a case, the criterion in (118) is still used because of the difficulty in computing the noncoherent detection metric for time-variant fading channels. The resulting performance degradation is now investigated.

The pairwise error probability P_d for a length-2 error event is

$$P_d = \mathbb{P}\left(\left\| \sum_{k=1}^{N} (h_k E_s + n'_k) \right\|^2 - \left\| \sum_{k=1}^{2} (\alpha_k h_k E_s + n''_k) + \sum_{k=3}^{N} (h_k E_s + n'_k) \right\|^2 \leq 0 \right). \qquad (138)$$

Define $\mathbf{a} = \left[\sum_{k=1}^{2} h_k E_s + n'_k, \sum_{k=1}^{2} \alpha_k h_k E_s + n''_k, \sum_{k=3}^{N} h_k E_s + n'_k \right]^T$. For DBPSK and MSK, $\alpha_1 = -\alpha_2$

and, therefore, the covariance $\mathbf{R}_a$ is simplified as[2]

$$\mathbf{R}_a = \begin{pmatrix} 2(1+\rho)E_s^2+2N_oE_s & 0 & A_1E_s^2 \\ 0 & 2\|\alpha_1\|^2(1-\rho)E_s^2+2N_oE_s & \alpha_1A_2E_s^2 \\ A_1E_s^2 & \alpha_1^*A_2E_s^2 & A_3E_s^2+(N-2)N_oE_s \end{pmatrix}, \quad (139)$$

where $A_1 = \frac{(\rho-\rho^{N-1})(1+\rho)}{1-\rho}$, $A_2 = \rho-\rho^{N-1}$ and $A_3 = \frac{2\rho(\rho^{N-2}-1)}{(1-\rho)^2} + \frac{(N-2)(1+\rho)}{1-\rho}$. The matrices $\mathbf{Q}_{1,2}$ are exactly the same as in (134). The detailed derivations of (139) and the eigenvalues of $(\mathbf{Q}_1 - \mathbf{Q}_2)\mathbf{R}_a$ are provided in Appendix F.

After straightforward but tedious manipulation, the pairwise error probability of DBPSK and MSK at high values of E_s/N_o can be approximated as

$$P_d^{\text{MSK}} \approx P_d^{\text{DBPSK}} \approx \frac{NN_o}{4(N-1)E_s} + \frac{N(1-\rho)}{6}, \quad (140)$$

where the first term converges to the error probability with coherent detection when N is sufficiently large. The second term is independent of E_s/N_o and represents an error floor at high E_s/N_o, while its effect deteriorates with increasing fading rate or observation width, N. Note that when deriving (140), we assume $N(1-\rho) \ll 1$, which is certainly true for practical noncoherent detectors and channel Doppler conditions. The proof of (140) is shown in Appendix C.

Equation (140) reveals some valuable insights on the optimization of a noncoherent detector under different channel conditions. For given channel characteristics, i.e., E_s/N_o and ρ, minimizing the pairwise error probability leads to a noncoherent detector with an optimized observation width

$$N_{\text{opt}} = \left[\sqrt{\frac{3N_o}{2(1-\rho)E_s}} + 1\right]_{\mathcal{R}} \quad (141)$$

where the operator $[\cdot]_{\mathcal{R}}$ rounds an argument to the nearest integer. When $E_s/N_o = 20$ dB and $\rho = 0.999$, based on (141) we can compute $N_{\text{opt}} = 5$; while for $E_s/N_o = 30$ dB and $\rho = 0.999$, $N_{\text{opt}} = 3$. These results imply that, in contrast to time-invariant fading channels, increasing the complexity of a noncoherent detector does not always guarantee a better

[2]In this subsection we limit our analysis to DBPSK and MSK in order to obtain the closed-form expression for the pairwise error probability. Other modulation schemes can be evaluated numerically.

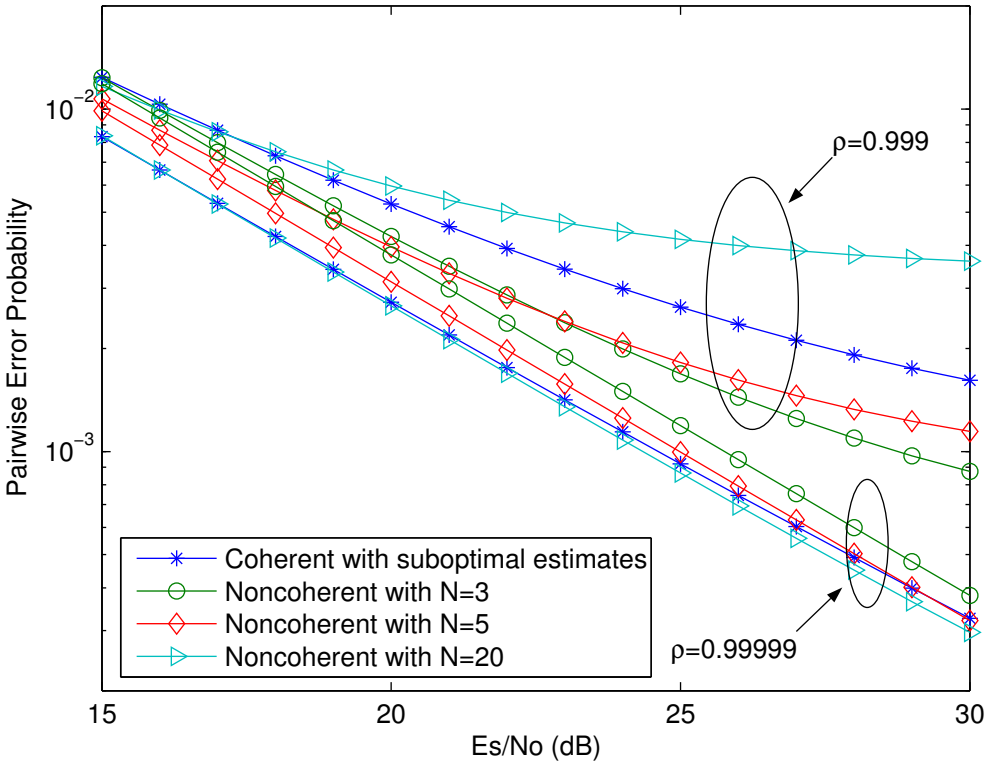


Figure 38: Performance of MSK noncoherent and coherent detection on time-variant fading channels.

detection performance on time-variant channels. Another interesting result is that, given two noncoherent detectors of different complexities, i.e., different observation widths, we can determine from (140) the channel conditions under which one is superior to the other. For example, given two detectors with $N=3$ and $N=5$, we can predict that at $\rho=0.999$ the detector with $N=3$ is superior when $E_s/N_o > 22.7\,\text{dB}$, and vice versa. On the other hand, if we fix $E_s/N_o=30\,\text{dB}$, the detector with $N=5$ is better when $\rho>0.9998$.

Fig. 38 shows the pairwise error probabilities of noncoherent MSK detectors having different complexities. Two typical values of ρ are considered, i.e., $\rho=0.99999$ and $\rho=0.999$, which correspond to slow and fast time-variant fading channels, respectively. When the channel is slowly time-varying, the second term in (140) is negligible at E_s/N_o values of practical interests and, thus, the performance of noncoherent detection improves with an increase in the observation width, i.e., the detector complexity. When N is sufficiently large,

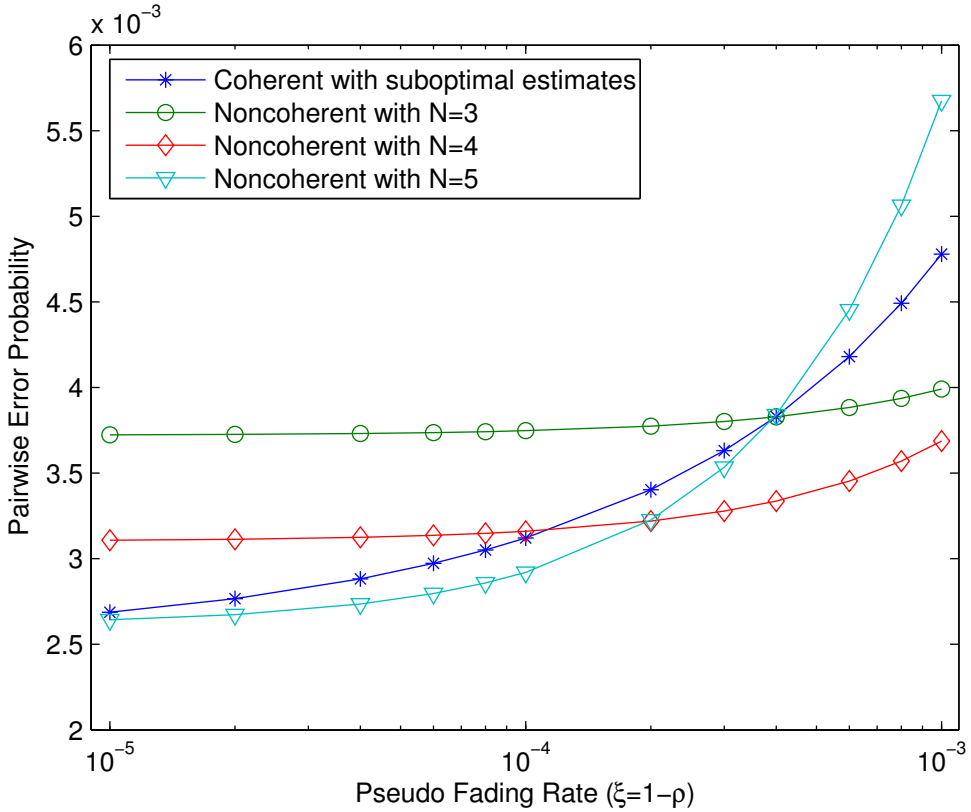


Figure 39: Performance of MSK noncoherent and coherent detection on time-variant fading channels with $E_s/N_o = 20$ dB.

e.g., $N = 20$, noncoherent detection outperforms coherent detection with imperfect channel knowledge in terms of error rate.

When the channel is rapidly time-varying, the second term in (140) begins to have an observable effect on the error probability with an increase in E_s/N_o. This results in better performance with a lower rather than higher complexity noncoherent detector at high E_s/N_o. As shown in Fig. 38, the noncoherent detector with $N = 3$ can achieve better performance as compared to the detector with $N = 5$, when E_s/N_o is greater than 23 dB. It is worth noting that both noncoherent detectors performs noticeably better than the coherent sequence detector using linear MMSE channel estimates, since the estimation performance is inevitably poor on fast fading channels. This observation raises an immediate question — is it possible to design a hybrid system that switches between noncoherent and coherent detectors based on the knowledge of channel characteristics, including E_s/N_o and Doppler

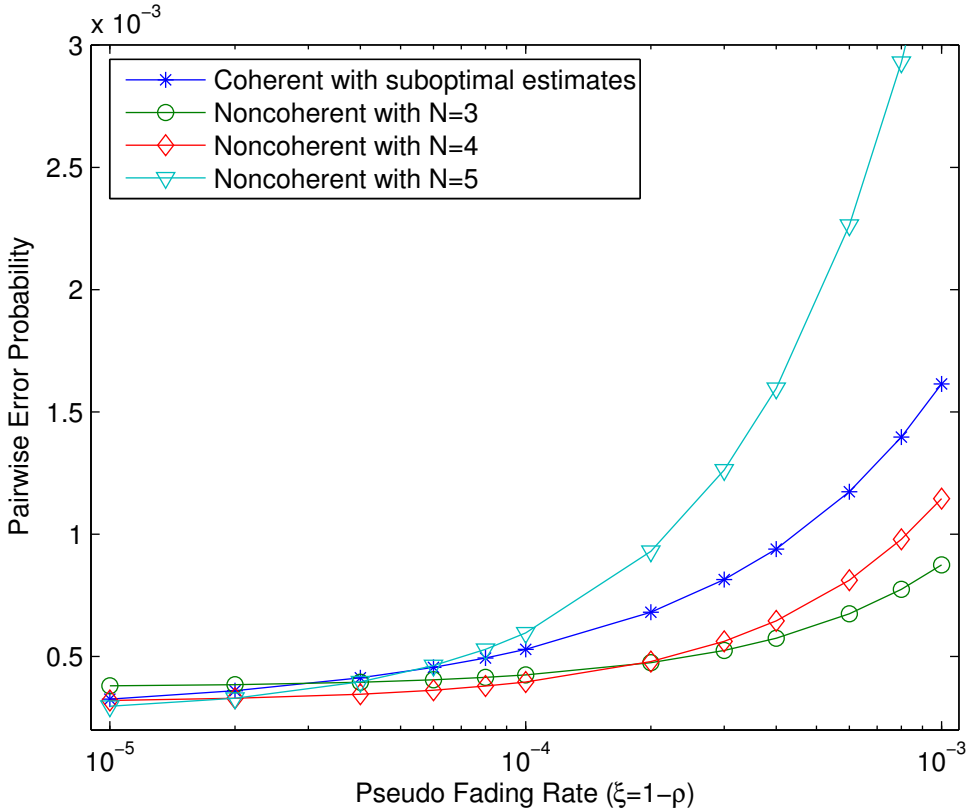


Figure 40: Performance of MSK noncoherent and coherent detection on time-variant fading channels with $E_s/N_o = 30$ dB.

(ρ), in order to optimize system overall performance in terms of service quality, receiver complexity and power efficiency.

6.3 A Hybrid Scheme on Time-variant Channels

The essential idea of a hybrid scheme is to use a coherent detector for time-invariant and slow time-variant channels, while using a noncoherent detector on relatively fast time-variant channels. However, the question remains as to when or in what conditions to trigger a detector switch, how to determine the optimal observation width for a noncoherent detector when the fading coefficient ρ itself is time-varying, and how much power gain can be achieved by using the hybrid system. Throughout this section we use MSK as an example and consider its pairwise error probability performance.

Let us start with a closer look at the error rate performance of different detectors.

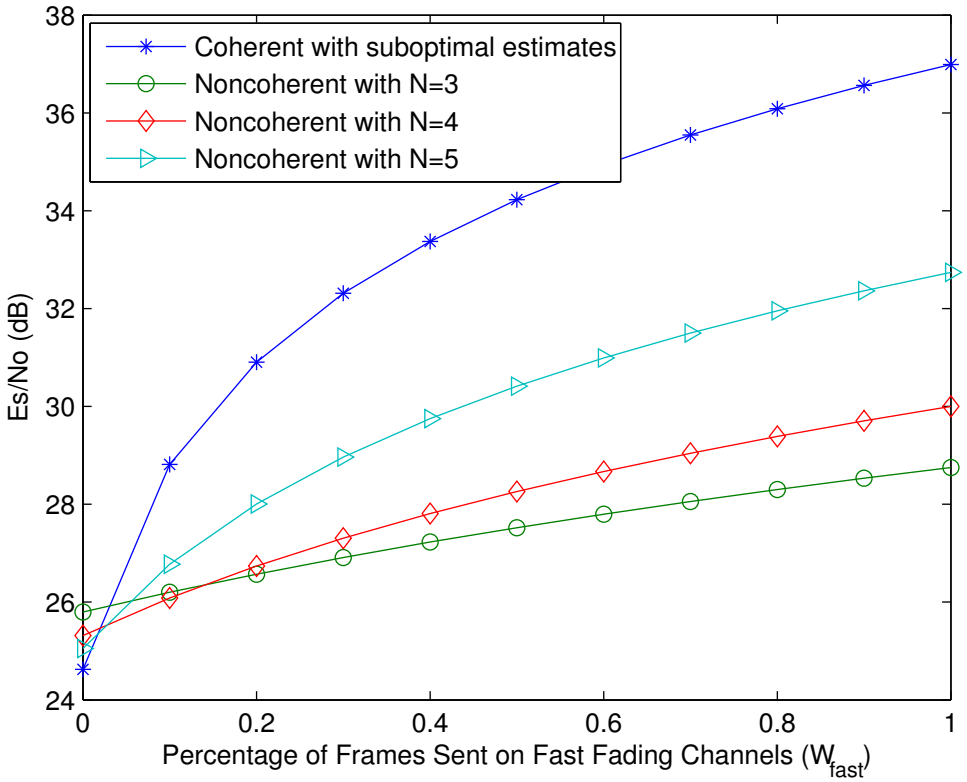


Figure 41: Power efficiency of noncoherent and coherent detection given service quality.

Instead of varying E_s/N_o while keeping ρ constant as in Fig. 38, in Figs. 39 and 40 we fix the value of E_s/N_o and vary the fading coefficient ρ over a wide range, e.g., $[0.999, 0.99999]$. For clarity, the abscissa in Figs. 39 and 40 indicates the variable $\xi = 1-\rho$, which is denoted as a *pseudo* fading rate, because a larger ξ indicates more rapid fading. In Fig. 39, E_s/N_o is set to 20 dB. We find that a less complex noncoherent detector is also less sensitive to variations in the fading rate. In particular, the error probability curve of the noncoherent detector with $N = 3$ is almost flat from $\xi = 10^{-5}$ to $\xi = 10^{-3}$. This feature is very valuable in practical system design, since no significant power boost is required to maintain service quality when the fading rate increases, which eases requirements on the linear operating range for RF devices[3]. In contrast, for coherent detection the transmitted power must be boosted. In Fig. 40, E_s/N_o is equal to 30 dB, and a similar phenomenon as in Fig. 39 is

[3]Different from the traditional concept of power control that is based on the observation of instantaneous or average channel gain, the power boost hereof is made to compensate the performance degradation owing to the channel variations given the fixed average channel gain.

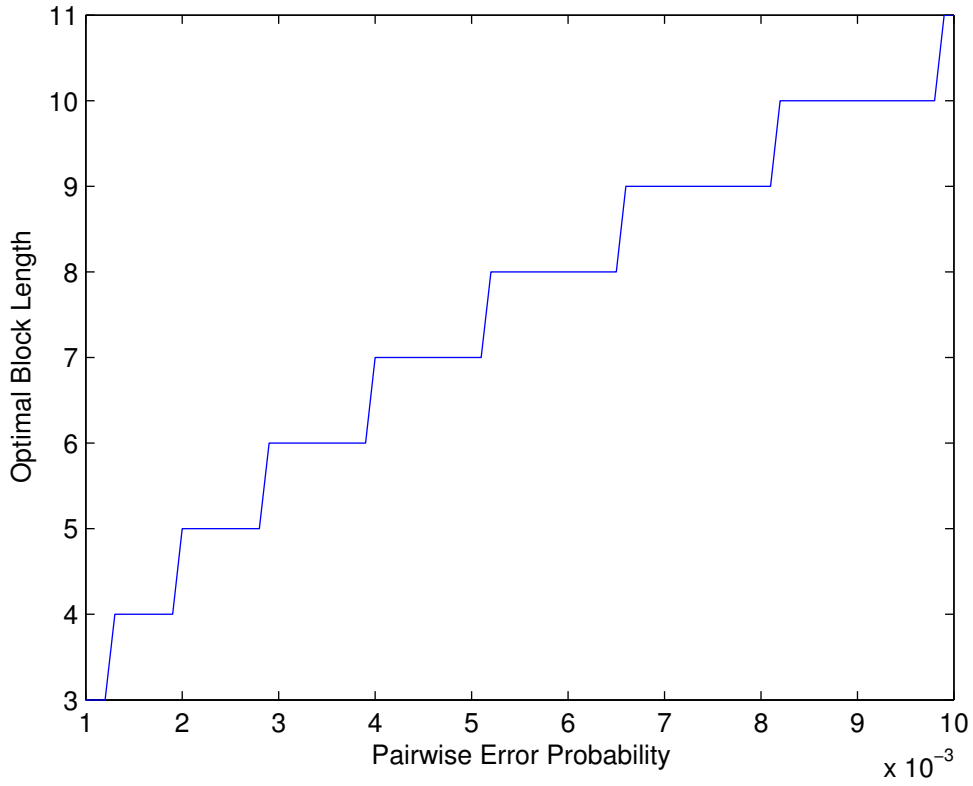


Figure 42: Optimal observation width of noncoherent detectors given service quality.

observed. Also, the noncoherent detector with $N = 3$ outperforms one with $N = 5$ when $\xi > 2\times10^{-4}$. From Figs. 39 and 40, we conclude that on fast fading channels it is preferable to use noncoherent detectors with $N \leq 5$ due to their moderate complexity and robustness to the variation of fading rate.

Now let us consider an artificial scenario where, during the transmission of a data frame, the channel exhibits either the *pseudo* fading rate $\xi = 10^{-5}$ or $\xi = 10^{-3}$. Over entire observation period, we let W_{fast} denote the percentage of data frames transmitted over the channel with $\xi = 10^{-3}$. We also assume that the transmitter is aware of the fading rate of the channel with zero-delay and can correspondingly adjust the transmitted power to maintain a constant error probability performance, i.e., service quality. Fig. 41 shows the average E_s/N_o required to maintain $P_d = 0.001$ for different detectors. When $W_{\text{fast}} < 0.03$, i.e., less than three percent of data frames are propagated over the channel with $\xi = 10^{-3}$, the coherent detector requires less power to maintain the service quality compared

to noncoherent detectors. When $W_{\text{fast}} > 0.03$, the power efficiency of the coherent detector significantly degrades. For example, the coherent detector requires 4 dB more power at $W_{\text{fast}} = 0.2$ and 7 dB more power at $W_{\text{fast}} = 0.5$ to maintain the same service quality as compared to a noncoherent detector with $N = 3$. This observation implies that coherent detection on fast time-variant fading channels is not power-efficient, while it performs well when $\xi \leq 10^{-5}$. As a result, we propose to use noncoherent detection when $\xi > 10^{-5}$ and use coherent detection when $\xi \leq 10^{-5}$.

In practice, the *pseudo* fading rate varies in time. We denote its profile as $\mathcal{P}(\xi)$, indicating the probability that a data frame is transmitted over a channel with *pseudo* fading rate ξ and $\int_0^1 \mathcal{P}(\xi)d\xi = 1$. For example, the profile of ξ in Fig. 41 is $W_{\text{fast}}\delta(\xi - 10^{-3}) + (1 - W_{\text{fast}})\delta(\xi - 10^{-5})$, where $\delta(\cdot)$ is the Delta function. Given the desired service quality, i.e., P_d^{QS}, and the trigger threshold ξ_{th}, the optimal observation width for noncoherent detection in terms of minimizing average transmitting power can be determined by

$$\begin{aligned} N_{\text{opt}} &= \min_N \int_{\xi_{\text{th}}}^1 \mathcal{E}(\xi, P_d^{\text{QS}}, N)\mathcal{P}(\xi)d\xi, \\ &= \min_N \int_{\xi_{\text{th}}}^1 \frac{6N}{N-1} \cdot \frac{\mathcal{P}(\xi)}{24P_d^{\text{QS}} - 4\,\xi N} d\xi, \end{aligned} \tag{142}$$

where $\mathcal{E}(\xi, P_d^{\text{QS}}, N)$ is the required transmitting power given ξ, P_d^{QS} and N, which can be computed from (140). In case that $\mathcal{P}(\xi)$ is not available, we simply assume ξ is uniformly distributed over the region of interests. Fig. 42 shows the optimal observation widths for different service qualities given $\mathcal{P}(\xi)$ uniformly distributed on the interval $(0, 0.001)$ and $\xi_{\text{th}} = 10^{-5}$. The result is surprising that the complexity of the optimal noncoherent detector reduces with increasing service quality. An intuitive explanation is that the noncoherent detectors are required to operate at relatively high E_s/N_o to meet the higher service qualities, conditions under which the effect of the error-floor term in (140) is salient and the detectors with a wider observation width may suffer from the early onset of an error floor.

CHAPTER VII

CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

This thesis addressed the problems in design and analysis of robust synchronization algorithms for uncoded and coded CPM systems. For completeness of the study on phase recovery, we also compared noncoherent and coherent detection on time-variant fading channels. The primary contributions and conclusions of the thesis are summarized as follows.

In Chapter II, using carefully selected nonsymmetric nonorthogonal exponential expansions for CPM signals, we proposed a joint timing and phase recovery method, whose acquisition performance is reliable on both time-invariant and time-variant channels, even with a short preamble. Moreover, in the proposed scheme, the complexity of the required front-end processor is significantly reduced without a salient sacrifice of accuracy, yielding a negligible performance loss in both synchronization and demodulation.

In Chapter III, a decision-directed parameter estimation method is presented in a form of combining linear least-squares estimation with MAP symbol detection in a PSP fashion. The resulting adaptive iterative algorithms, i.e., A-SISO algorithms, are applied to SCCPM systems to implement joint phase synchronization and CPM demodulation. We noticed that CPM can be viewed as a subclass of RIC codes. Given a RIC trellis, it is possible to reduce or remove the effects of cycle slip, false lock and hangup. However, for SCCPM systems, especially at low SNRs, when the *bidirectional* A-SISO algorithms try to take advantage of rotational invariance during phase estimation, the MAP-based CPM demodulation procedure suffers a performance loss caused also by the rotational invariance. Fortunately, by using the *forward-only* A-SISO algorithms, we can achieve a robust performance in both synchronization and demodulation. When SNR is high and the carrier phase is time-invariant, phase rotations rarely occur and, thus, CPM systems act as non-RIC

codes. In this case, the *bidirectional* A-SISO algorithms outperform the *forward-only* A-SISO algorithms, and the A-SISO algorithms using the smoothed phase estimates achieves a near-optimal performance in both estimation and detection.

In Chapter IV, we further investigated the convergence characteristics of iterative phase synchronization and detection for rotationally invariant turbo-like codes, including SCCPM. Based on a semi-theoretical analysis, we show that the occurrence of cycle slips could result in convergence oscillation during iterative detection, and more important, the convergence performance of SEP-based A-SISO algorithms is more robust than that of the PSP-based counterparts, when tracking a fast time-variant carrier phase.

Finally, a turbo synchronizer is devised in Chapter V for joint time and phase estimation at iterative SCCPM receiver, by combining the noniterative synchronizer proposed in Chapter II with the adaptive iterative structure proposed in Chapter III. Simulation results show that the proposed synchronizer excels in both acquiring and tracking the time and phase errors at low SNRs and it is robust even when incorrect channel modeling is present. However, owing to the use of preamble and pilot symbols to prevent false locks, a significant power loss is introduced.

In this thesis, we also computed the error probabilities of noncoherent detection in Chapter VI for full-response CPM and DPSK, which clearly indicates how noncoherent detection performance is affected by detector complexity, fading rate and SNR. We concluded that on fast fading channels, noncoherent detection with properly selected block lengths outperforms coherent detection with practical channel estimation. Furthermore, a less rather than more complex noncoherent detector might achieve a better error rate performance on fast fading channels. To achieve an optimal system performance, we designed a hybrid system, in which a coherent detector is activated on time-invariant and slow fading channels, while a noncoherent detector is used when channels are fast time-varying.

7.2 Future Work

The future work can proceed along the several lines as follows.

Within a turbo synchronizer for joint time and phase recovery, to prevent false locks, we use a preamble to acquire the unknown parameters and also use pilot symbols to assist decision-directed parameter tracking during iterative detection. However, the use of preamble and pilots introduces a significant power loss. The problems remain open on how to design a preamble sequence, how to place pilots, and how to distribute the power among the preamble, pilots and data, in order for achieving optimal power efficiency.

Turbo-like codes with rotational invariance have been investigated in this thesis through simulations and theoretical analysis. It is observed that when unreliable phase synchronization is inevitable because of low operating SNR and high phase dynamics, these codes still can achieve robust detection performance for well-designed receivers. Low density parity check (LDPC) codes are another set of capacity-approaching codes, and are expected to be resilient to strong phase noise. Therefore, the design and analysis of LDPC codes with phase rotational invariance is of great practical interests, including code and constellation design, development of complexity-reduced detection algorithms and convergence analysis of iterative receivers.

We have studied the error probabilities of noncoherent detection for full-response CPM. The extension to partial-response CPM is apparently necessary and valuable, although it is not trivial. Moreover, the information-theoretic analysis, including both cutoff rate and channel capacity, for noncoherent detection on fading channels is an open area and has been paid little attention in the literature. This analysis could help understand the principle of noncoherent data transmission, and construct new schemes combining noncoherent detection with coherent detection to maximize an achievable data rate on practical fading channels.

APPENDIX A

SMOOTHING ESTIMATION

The time-variant state-space model with state-vector $\boldsymbol{\theta}_i$ can be modelled as follows,

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{F}_i\boldsymbol{\theta}_i + \boldsymbol{G}_i\boldsymbol{w}_i, \tag{A.1a}$$

$$\mathbf{y}_i = \boldsymbol{H}_i\boldsymbol{\theta}_i + \boldsymbol{v}_i, \;\; 1 \leq i \leq N_u \tag{A.1b}$$

where $\boldsymbol{F}_i$, $\boldsymbol{G}_i$ and $\boldsymbol{H}_i$ are time-variant coefficient matrices with dimension $L_\theta \times L_\theta$, $L_\theta \times L_\theta$ and $J \times L_\theta$, respectively, $\boldsymbol{w}_i$ and $\boldsymbol{v}_i$ are zero-mean Gaussian-distributed random vectors with dimension $L_\theta \times 1$ and $J \times 1$, respectively. Note that $\boldsymbol{w}_i$ is uncorrelated with $\boldsymbol{v}_i$ and their covariance matrices are $\boldsymbol{W}_i$ and $\boldsymbol{V}_i$. Without loss of generality, we assume that $\boldsymbol{\theta}_i$ is a zero-mean random vector with dimension $L_\theta \times 1$ and covariance matrix $\boldsymbol{\xi}_i$.

Using the two-filter smoothing formulas in [90], the smoothed estimate of $\boldsymbol{\theta}_i$ is expressed as

$$\hat{\boldsymbol{\theta}}_{i|N_u} = \boldsymbol{C}_{i|N_u}(\boldsymbol{C}_{i|i}^{-1}\hat{\boldsymbol{\theta}}_{i|i} + (\boldsymbol{C}_{i|i+1}^{b})^{-1}\hat{\boldsymbol{\theta}}_{i|i+1}^{b}) \tag{A.2}$$

with $\boldsymbol{C}_{i|N_u}^{-1} = \boldsymbol{C}_{i|i}^{-1} + (\boldsymbol{C}_{i|i+1}^{b})^{-1} - \boldsymbol{\xi}_i^{-1}$. If no *a priori* information is available, then $\boldsymbol{\xi}_i^{-1} = \mathbf{0}$. Using the matrix inversion lemma [90], we can rewrite (A.2) as

$$\hat{\boldsymbol{\theta}}_{i|N_u} = \hat{\boldsymbol{\theta}}_{i|i} - \boldsymbol{C}_{i|i}(\boldsymbol{C}_{i|i} + \boldsymbol{C}_{i|i+1}^{b})^{-1}(\hat{\boldsymbol{\theta}}_{i|i} - \hat{\boldsymbol{\theta}}_{i|i+1}^{b}). \tag{A.3}$$

When unknown parameters are deterministic, the state-space model in (A.1) is still valid by assuming that the process noise $\boldsymbol{w}_i$ vanishes. It has been shown in [103] that the adaptive recursive-least-squares (RLS) algorithm is equivalent to the Kalman recursion by letting $\boldsymbol{F}_i = \lambda \cdot \boldsymbol{I}$, where λ is the forgetting factor in the RLS algorithm. Moreover, let $\lambda = 1$ and let the Kalman gain be $\boldsymbol{K}_{f,i} = \eta \boldsymbol{H}_i^T$. Then the forward Kalman filtering reduces to the LMS algorithm, viz.,

$$\hat{\boldsymbol{\theta}}_{i+1|i+1} = \hat{\boldsymbol{\theta}}_{i|i} + \eta \boldsymbol{H}_{i+1}^T(y_{i+1} - \boldsymbol{H}_{i+1}\hat{\boldsymbol{\theta}}_{i|i}), \tag{A.4}$$

where η is the stepsize. If we further assume that $\boldsymbol{H}_i$ and $\boldsymbol{V}_i$ are fixed during the entire observation, the inverses of covariance matrices $\boldsymbol{C}_{i|i}$ and $\boldsymbol{C}^b_{i|i+1}$ can be simplified as

$$\begin{aligned} \boldsymbol{C}^{-1}_{i|i} &= i\boldsymbol{H}^T\boldsymbol{V}^{-1}\boldsymbol{H}, \\ (\boldsymbol{C}^b_{i|i+1})^{-1} &= (N_u - i)\boldsymbol{H}^T\boldsymbol{V}^{-1}\boldsymbol{H}. \end{aligned} \tag{A.5}$$

Substituting (A.5) into (A.3), the smoothed estimate in this case is given by

$$\hat{\boldsymbol{\theta}}_{i|N_u} = \frac{i}{N_u}\hat{\boldsymbol{\theta}}_{i|i} + \frac{N_u - i}{N_u}\hat{\boldsymbol{\theta}}^b_{i|i+1}, \tag{A.6}$$

which implies that the smoothed estimate $\hat{\boldsymbol{\theta}}_{i|N_u}$ is a weighted average of $\hat{\boldsymbol{\theta}}_{i|i}$ and $\hat{\boldsymbol{\theta}}^b_i$, where the weighting factors represent the reliability of the corresponding forward and backward estimates.

APPENDIX B

EQUIVALENCE OF (59) AND (60)

With the state-space model in (A.1), evaluating the APP of a hypothesized sequence is equivalent to evaluating the following least-squares,

$$\Lambda(\tilde{\mathbf{u}}_1^{N_u}) = \sum_{i=1}^{N_u} \parallel \mathbf{y}_i - \boldsymbol{H_i}\hat{\boldsymbol{\theta}}_i \parallel^2 . \tag{B.1}$$

By assuming $\boldsymbol{\theta}_i$ is fixed over the entire observation, the smoothed estimate of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}}_s = \boldsymbol{B}^{-1}\boldsymbol{z} \tag{B.2}$$

with $\boldsymbol{B} = \sum_{i=1}^{N_u} \boldsymbol{H}_i^H \boldsymbol{H}_i$ and $\boldsymbol{z} = \sum_{i=1}^{N_u} \boldsymbol{H}_i^H \mathbf{y}_i$. Similarly, the forward and backward predicted estimates of $\boldsymbol{\theta}$ are

$$\hat{\boldsymbol{\theta}}_k^f = (\boldsymbol{B}_k^f)^{-1}\boldsymbol{z}_k^f, \text{ and } \hat{\boldsymbol{\theta}}_k^b = (\boldsymbol{B}_k^b)^{-1}\boldsymbol{z}_k^b, \tag{B.3}$$

respectively, with $\boldsymbol{B}_k^f = \sum_{i=1}^{k} \boldsymbol{H}_i^H \boldsymbol{H}_i$, $\boldsymbol{B}_k^b = \sum_{i=k}^{N_u} \boldsymbol{H}_i^H \boldsymbol{H}_i$, $\boldsymbol{z}_k^f = \sum_{i=1}^{k} \boldsymbol{H}_i^H \mathbf{y}_i$ and $\boldsymbol{z}_k^b = \sum_{i=k}^{N_u} \boldsymbol{H}_i^H \mathbf{y}_i$. For simplicity, we assume $\boldsymbol{B}$, $\boldsymbol{B}_k^f$ and $\boldsymbol{B}_k^b$ are invertible. In the light of (57), we rewrite $\hat{\boldsymbol{\theta}}_s$ as

$$\hat{\boldsymbol{\theta}}_s = \hat{\boldsymbol{\theta}}_k^f + \hat{\boldsymbol{\theta}}_{k+1}^b - \hat{\boldsymbol{\theta}}_c, \tag{B.4}$$

where $\hat{\boldsymbol{\theta}}_c$ is a correction term given by

$$\hat{\boldsymbol{\theta}}_c = (\boldsymbol{B}_k^f)^{-1}\left[(\boldsymbol{B}_k^f)^{-1} + (\boldsymbol{B}_{k+1}^b)^{-1}\right]^{-1}\hat{\boldsymbol{\theta}}_k^f + (\boldsymbol{B}_{k+1}^b)^{-1}\left[(\boldsymbol{B}_k^f)^{-1} + (\boldsymbol{B}_{k+1}^b)^{-1}\right]^{-1}\hat{\boldsymbol{\theta}}_{k+1}^b. \tag{B.5}$$

Finally, we express the optimal least-squares under the entire observation as

$$\Lambda_s(\tilde{\mathbf{u}}_1^{N_u}) = \sum_{i=1}^{N_u} (\mathbf{y}_i - \boldsymbol{H_i}\hat{\boldsymbol{\theta}}_s)^H \mathbf{y}_i, \tag{B.6a}$$

$$= \sum_{i=1}^{k} (\mathbf{y}_i - \boldsymbol{H_i}\hat{\boldsymbol{\theta}}_s)^H \mathbf{y}_i + \sum_{i=k+1}^{N_u} (\mathbf{y}_i - \boldsymbol{H_i}\hat{\boldsymbol{\theta}}_s)^H \mathbf{y}_i \tag{B.6b}$$

$$= \sum_{i=1}^{k} (\mathbf{y}_i - \boldsymbol{H_i}\hat{\boldsymbol{\theta}}_k^f)^H \mathbf{y}_i + \sum_{i=k+1}^{N_u} (\mathbf{y}_i - \boldsymbol{H_i}\hat{\boldsymbol{\theta}}_{k+1}^b)^H \mathbf{y}_i + (\boldsymbol{z}_k^f)^H(\hat{\boldsymbol{\theta}}_k^f - \hat{\boldsymbol{\theta}}_s) + (\boldsymbol{z}_{k+1}^b)^H(\hat{\boldsymbol{\theta}}_{k+1}^b - \hat{\boldsymbol{\theta}}_s) \quad \text{(B.6c)}$$

$$= \Lambda_f(\tilde{\mathbf{u}}_1^k) + \Lambda_b(\tilde{\mathbf{u}}_{k+1}^{N_u}) + \Lambda_k^c, \quad \text{(B.6d)}$$

where the first term in (B.6c) denoted as $\Lambda_f(\tilde{\mathbf{u}}_1^k)$ is the the minimized least-squares with the present and past observation, the second term in (B.6c) denoted as $\Lambda_b(\tilde{\mathbf{u}}_{k+1}^{N_u})$ is the minimized least-squares with the future signal record, and the remaining part in (B.6c) is defined as Λ_k^c. After some algebra, it can be shown that

$$\Lambda_k^c = (\hat{\boldsymbol{\theta}}_{k+1}^b - \hat{\boldsymbol{\theta}}_c)^H(\boldsymbol{B}_k^f)(\hat{\boldsymbol{\theta}}_{k+1}^b - \hat{\boldsymbol{\theta}}_c) + (\hat{\boldsymbol{\theta}}_k^f - \hat{\boldsymbol{\theta}}_c)^H(\boldsymbol{B}_{k+1}^b)(\hat{\boldsymbol{\theta}}_k^f - \hat{\boldsymbol{\theta}}_c), \quad \text{(B.7)}$$

which is the exact same as the binding factor defined in [99].

APPENDIX C

PROOF OF THE EQUATION (62)

Define the forward survivor at state s_k as the symbol sequence with the maximum APP among all symbol sequences ending with state s_k. Let $\hat{\boldsymbol{\theta}}_k^{\mathcal{S}_\mathcal{F}}(s_k)$ denote the *forward survived* LLSE of $\boldsymbol{\theta}_k$ at state s_k, which can be obtained by running the forward linear estimator along the forward survivor ending with state s_k. Similarly, the *backward survived* LLSE of $\boldsymbol{\theta}_k$ associated with the state s_{k+1} is defined as $\hat{\boldsymbol{\theta}}_k^{\mathcal{S}_\mathcal{B}}(s_{k+1})$. Moreover, we also denote $\hat{\boldsymbol{\theta}}_k^{\mathcal{B}}(t_k)$ as the *branch* estimate of $\boldsymbol{\theta}_k$, which is used to update the metric of the branch t_k connecting s_k and s_{k+1}. Let $\hat{\boldsymbol{\theta}}_k^{\mathcal{B}}$ be the collection of all *branch* estimates at the k-th time epoch. Note that $\hat{\boldsymbol{\theta}}_k^{\mathcal{S}_\mathcal{F}}(s_k)$ and $\hat{\boldsymbol{\theta}}_k^{\mathcal{B}}(t_k)$ are not necessarily same, and neither are $\hat{\boldsymbol{\theta}}_k^{\mathcal{S}_\mathcal{B}}(s_k)$ and $\hat{\boldsymbol{\theta}}_k^{\mathcal{B}}(t_k)$. The *branch* estimate $\hat{\boldsymbol{\theta}}_k^{\mathcal{B}}(t_k)$ could be a refined version of $\hat{\boldsymbol{\theta}}_k^{\mathcal{S}_\mathcal{F}}(s_k)$ and $\hat{\boldsymbol{\theta}}_k^{\mathcal{S}_\mathcal{F}}(s_k)$ by considering additional information extracted from the past/future observation. Of course, setting them equal yields a simplified implementation but a performance degradation as well. Finally, with above notation, the forward state metric can be computed as

$$
\begin{aligned}
\alpha(s_{k+1}) &= \sum_{\tilde{\mathbf{u}}_1^k : s_{k+1}} p\left(\mathbf{y}_1^k | \tilde{\mathbf{u}}_1^k, \hat{\boldsymbol{\theta}}_1^k(\mathbf{y}_1^k, \tilde{\mathbf{u}}_1^k)\right) \mathbb{P}\left(\tilde{\mathbf{u}}_1^k\right) \\
&= \sum_{s_k \in \mathcal{S}} \sum_{\tilde{\mathbf{u}}_1^{k-1} : s_k} p\left(\mathbf{y}_1^{k-1} | \tilde{\mathbf{u}}_1^{k-1}, \hat{\boldsymbol{\theta}}_1^{k-1}\right) p\left(\mathbf{y}_k | s_k, s_{k+1}, \hat{\boldsymbol{\theta}}_k\right) \mathbb{P}\left(\tilde{\mathbf{u}}_1^k\right) \\
&\approx \sum_{s_k \in \mathcal{S}} \left\{ \sum_{\tilde{\mathbf{u}}_1^{k-1} : s_k} p\left(\mathbf{y}_1^{k-1} | \tilde{\mathbf{u}}_1^{k-1}, \hat{\boldsymbol{\theta}}_1^{k-1}\right) \mathbb{P}\left(\tilde{\mathbf{u}}_1^{k-1}\right) \right\} p\left(\mathbf{y}_k | s_k, s_{k+1}, \hat{\boldsymbol{\theta}}_k^{\mathcal{B}}(t_k)\right) \mathbb{P}\left(s_{k+1} | s_k\right) \qquad \text{(C.1a)} \\
&= \sum_{s_k \in \mathcal{S}} \alpha(s_k) \gamma(s_k, s_{k+1}), \qquad \text{(C.1b)}
\end{aligned}
$$

where in (C.1a) we approximate $\hat{\boldsymbol{\theta}}_k$ with the *branch* estimate $\hat{\boldsymbol{\theta}}_k^{\mathcal{B}}(t_k)$, and in (C.1b) we denote the branch metric as $\gamma(s_k, s_{k+1}) = p\left(\mathbf{y}_k | s_k, s_{k+1}, \hat{\boldsymbol{\theta}}_k^{\mathcal{B}}(t_k)\right) \mathbb{P}\left(s_{k+1} | s_k\right)$. A similar derivation can be shown for the second inner summation term in (59).

The remaining unsolved question is how to update survived estimates efficiently. Recall that there are M branches starting from and merging into each state. Hence there are

M candidate estimates competing for the *forward survived* estimate at each state. The criterion for the selection of $\hat{\boldsymbol{\theta}}_k^{\mathcal{S}_\mathcal{F}}(s_k)$, is

$$t_{k-1}^{\mathcal{S}_\mathcal{F}}(s_k) = \arg \max_{t_k : s_{k-1} \to s_k} \alpha(s_{k-1})\gamma(s_{k-1}, s_k), \tag{C.2a}$$

$$\hat{\boldsymbol{\theta}}_k^{\mathcal{S}_\mathcal{F}}(s_k) = \Upsilon\big(\hat{\boldsymbol{\theta}}_{k-1}^{\mathcal{S}_\mathcal{F}}(s_{k-1}, \mathbf{y}_{k-1}, t_{k-1}^{\mathcal{S}_\mathcal{F}}(s_k))\big), \tag{C.2b}$$

where $t_{k-1}^{\mathcal{S}_\mathcal{F}}(s_k)$ is the *forward survived* branch ending with the state s_k, and $\Upsilon(\cdot)$ is the parameter updating function. The selection of $\hat{\boldsymbol{\theta}}_k^{\mathcal{S}_\mathcal{B}}(s_k)$ is performed in a similar way. Note that in (C.2) a single survived branch and, thus, a single survivor is chosen at each state. However, it is not necessary to retain only one survivor at each state. With multiple survivors at each state, a more reliable estimate can be achieved. However, this also causes an unwanted complexity increase. The tradeoff between performance and complexity needs further investigation in the future.

APPENDIX D

PROOF OF THE EQUATION (127)

With imperfect channel knowledge, the error probability of a two-error event on time-invariant channels is

$$\begin{aligned} P_d &= \mathbb{P}\left(\|y_1-\hat{h}f_1\|^2+\|y_2-\hat{h}f_2\|^2 \leq \|y_1-\hat{h}x_1\|^2+\|y_2-\hat{h}x_2\|^2\right), \\ &= \mathbb{P}\left(\|\hat{h}_1 D_1+u_1\|^2+\|\hat{h}_2 D_2+u_2\|^2 \leq \|u_1\|^2+\|u_2\|^2\right), \\ &= \mathbb{P}\left(\|D_1\|^2+\|D_2\|^2 \leq -2\mathrm{Re}\{\langle D_1,u_1\rangle+\langle D_2,u_2\rangle\}\right), \end{aligned} \tag{D.1}$$

where $u_i = (h-\hat{h})x_i+n_i$ and we assume $\|\hat{h}\|^2 = 1-P_h \approx 1$. Since u_1 and u_2 are correlated, the variance of the term $w = \mathrm{Re}\{\langle D_1,u_1\rangle+\langle D_2,u_2\rangle\}$ is

$$\mathrm{Var}(w) = 2(\|D_1\|^2+\|D_2\|^2)(N_o+P_hE_s)+2\mathrm{Re}\{\langle D_1,D_2\rangle\}P_hE_s, \tag{D.2}$$

and therefore K_h in the equation (127) is $\frac{2\mathrm{Re}\{\langle D_1,D_2\rangle\}}{\|D_1\|^2+\|D_2\|^2}$.

APPENDIX E

PROOF OF THE EQUATION (128)

We assume the linear MMSE channel estimation is used. Given the first-order Gauss-Markov channel model in (111), the the variance of estimation error at the i-th time epoch can be computed recursively by the Riccati equation as following,

$$P_{h,i+1} = \rho^2 P_{h,i} + (1-\rho^2) - \frac{\rho^2 P_{h,i} E_s}{N_o + P_{h,i} E_s}. \tag{E.1}$$

Let the time index i go to infinity and assume $P_{h,i}$ converges to P_h. Then the equation (E.1) can be rewritten as

$$P_h = \rho^2 P_h + (1-\rho^2) - \frac{\rho^2 P_h E_s}{N_o + P_h E_s}, \tag{E.2}$$

from which P_h can be easily obtained as in (128).

APPENDIX F

PROOF OF THE EQUATIONS (139) AND (140)

Let $r_{i,j}$ be the element at the i-th row and j-th column of the matrix $\mathbf{R}_a$. The computation of $r_{1,1}$, $r_{2,1}$ and $r_{2,2}$ is straightforward and, thus, here we only brief the derivation of $r_{3,1}$, $r_{3,2}$ and $r_{3,3}$,

$$\begin{aligned} r_{3,1} &= \mathbb{E}\left\{\left(\sum_{k=3}^{N} h_k E_s + n'_k\right)\left(\sum_{k=1}^{2} h_k E_s + n'_k\right)^*\right\}, \\ &= \sum_{k=1}^{N-2}\left(\rho^k + \rho^{k+1}\right) E_s^2, \\ &= A_1 E_s^2, \end{aligned} \tag{F.1a}$$

$$\begin{aligned} r_{3,1} &= \mathbb{E}\left\{\left(\sum_{k=3}^{N} h_k E_s + n'_k\right)\left(\sum_{k=1}^{2} \alpha_k h_k E_s + n''_k\right)^*\right\}, \\ &= \sum_{k=1}^{N-2}\left(\alpha_2^* \rho^k + \alpha_1^* \rho^{k+1}\right) E_s^2, \\ &= \alpha_1^* \rho^{N-1} E_s^2 + \alpha_2^* \rho E_s^2 + \sum_{k=2}^{N-2}\left(\alpha_1^* + \alpha_2^*\right)\rho^k E_s^2, \\ &= \alpha_1^* A_2 E_s^2, \end{aligned} \tag{F.1b}$$

$$\begin{aligned} r_{3,1} &= \mathbb{E}\left\{\left(\sum_{k=3}^{N} h_k E_s + n'_k\right)\left(\sum_{k=3}^{N} h_k E_s + n'_k\right)^*\right\}, \\ &= \sum_{k=3}^{N}\sum_{m=3}^{N} \rho^{|k-m|} E_s^2 + (N-2) E_s N_o, \\ &= A_3 E_s^2 + (N-2) E_s N_o. \end{aligned} \tag{F.1c}$$

Let $\rho = 1 - \xi$ and assume $\xi \ll 1$ for fading channels of practical interests. Moreover, by using Taylor's expansion, $A_{1,2,3}$ can be further approximated as

$$A_1 \approx 2(N-2) - N(N-2)\xi, \tag{F.2a}$$

$$A_2 \approx -(N-2)\xi, \tag{F.2b}$$

$$A_3 \approx (N-2)^2 - \frac{1}{3}(N-1)(N-2)(N-3)\xi. \tag{F.2c}$$

The eigenvalues of $(\mathbf{Q}_1 - \mathbf{Q}_2)\mathbf{R}_a$ are derived as

$$\frac{\lambda_{1,2}}{E_s} = B_1 E_s \pm \sqrt{B_2 E_s^2 + B_3 E_s N_o + 4(N-1)N_o^2}, \quad \text{and} \quad \lambda_3 = 0, \tag{F.3}$$

where $B_{1,2,3}$ are given by

$$B_1 \approx 2(N-1) - N(N-2)\xi - (1 + \|\alpha_1\|^2)\xi, \tag{F.4a}$$

$$B_2 \approx 4(N-1)^2 + \mathcal{O}(\xi), \tag{F.4b}$$

$$B_3 \approx 4N(N-1) - 2N(1 - \|a\|^2)\xi - \frac{4}{3}N(N-1)(N-2)\xi. \tag{F.4c}$$

At high SNRs, the two non-zero eigenvalues can be approximated as

$$\frac{\lambda_{1,2}}{E_s} = B_1 E_s \pm \left(\sqrt{B_2} E_s + \frac{B_3 N_o}{2\sqrt{B_2}}\right). \tag{F.5}$$

Using *Proposition* 2 again, the error probability P_d is computed as

$$\begin{aligned} P_d &= \frac{\lambda_2}{\lambda_2 - \lambda_1}, \\ &\approx \frac{(\sqrt{B_2} - B_1)E_s + \frac{B_3 N_o}{2\sqrt{B_2}}}{2\sqrt{B_2}E_s + \frac{B_3 N_o}{\sqrt{B_2}}}, \\ &= \frac{1}{2} + \frac{B_3 N_o}{4B_2 E_s} - \frac{B_1}{2\sqrt{B_2}}, \\ &\approx \frac{N N_o}{4(N-1)E_s} + \frac{N\xi}{6}\square \end{aligned} \tag{F.6}$$

REFERENCES

[1] T. Aulin and C.-E. Sundberg,"Continuous phase modulation-Part I: Full response signaling," *IEEE Trans. Commun.*, vol. 29, pp. 196-209, March 1981.

[2] T. Aulin, N. Rydbeck and C.-E. Sundberg,"Continuous phase modulation-Part II: Partial response signaling," *IEEE Trans. Commun.*, vol. 29, pp. 210-225, March 1981.

[3] T. Aulin,"Symbol error probability bounds for coherently viterbi detected continuous phase modulated signals," *IEEE Trans. Commun.*, vol. 29, pp. 1707-1715, Nov. 1981.

[4] J. B. Anderson, T. Aulin and C. E. Sundberg, *Digital Phase Modulation*. New York: Plenum Press, 1986.

[5] V. A. Dubendorf, *Wireless Data Technologies Reference Handbook*. New Jersey: John Wiley&Sons, 2003.

[6] U. Mengali and A. N. D'Andrea, *Synchronization techniques for digital receivers*. New York: Plenum Press, 1997.

[7] K. R. Narayanan and G. L. Stüber, "A serial concatenation approach to iterative demodulation and decoding," *IEEE Trans. Commun.*, vol. 47, pp. 956-961, July 1999.

[8] P. Moqvist and T. M. Aulin, "Serially concatenated continuous phase modulation with iterative decoding," *IEEE Trans. Commun.*, vol. 49, pp. 1901-1915, Nov. 2001.

[9] J. L. Stensby, *Phase-Locked Loops: Theory and Applications*. Boca Raton: CRC Press, 1997.

[10] A. N. D'Andrea, U. Mengali and R. Morelli, "A digital appraoch to clock recovery in generalized minimum shift keying," *IEEE Veh. Tech.*, vol. 39, pp. 227-234, Aug. 1990.

[11] J. H. Lodge and M. Moher, "Maximum likelihood sequence estimation of CPM signals transmitted over Rayleigh flat-fading Channels," *IEEE Trans. Commun.*, vol. 38, pp. 787-794, June 1990.

[12] G. M. Vitetta, U. Mengali and D. P. Taylor, "Blind detection of CPM signals transmitted over frequency-flat fading channels," *IEEE Trans. Veh. Tech.*, vol. 47, pp. 961-968, Aug. 1998.

[13] A. N. D'Andrea, U. Mengali and R. Morelli, "Symbol timing estimation with CPM modulation," *IEEE Trans. Commun.*, vol. 44, pp. 1362-1372, Oct. 1996.

[14] F. Patenaude, J. H. Lodge and P. A.Galko, "Symbol timing tracking for continuous phase modulation over fast flat-fading channels," *IEEE Trans. Veh. Tech.*, vol. 40, pp. 615-626, Aug. 1991.

[15] A. Ginesi, U. Mengali and M. Morelli, "Symbol and superbaud timing recovery in multi-h continuous-phase modulation," *IEEE Trans. Commun.*, vol. 47, pp. 664-666, May 1999.

[16] R. Mehlan, Y.-E. Chen and H. Meyr, "A fully digital feedforward MSK demodulator with joint frequency offset and symbol timing estimation for burst mode mobile radio," *IEEE Trans. Veh. Tech.*, vol. 42, pp. 434-443, Nov. 1993.

[17] U. Lambrette and H. Meyr, "Two Timing Recovery Algorithms for MSK," *Proc. IEEE ICC'94*, New Orleans, pp. 1155-1159, May 1994.

[18] J. Huber and W. Liu, "Data-aided synchronization of coherent CPM-receivers," *IEEE Trans. Commun.*, vol. 40, pp. 178-188, Jan. 1992.

[19] W. Tang and E. Shwedyk, "ML Estimation of Symbol Timing and Carrier Phase for CPM in Walsh Signal Space," *IEEE Trans. Commun.*, vol. 49, pp. 969-974, June 2001.

[20] M. Morelli, U. Mengali and G. M. Vitetta, "Joint phase and timing recovery with CPM signals," *IEEE Trans. Commun.*, vol. 45, pp. 867-876, July 1997.

[21] R. Morelli and G. M. Vitetta, "Joint phase and timing recovery for MSK-type signals," *IEEE Trans. Commun.*, vol. 48, pp. 1997-1999, Dec. 2000.

[22] G. K. Kaleh, "Simple coherent receivers for partial response continuous phase modulation," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 1427-1436, Dec. 1989.

[23] S. J. Simmons and P. H. Wittke,"Low complexity decoders for constant envelope modulations," *IEEE Trans. Commun.*, vol. 31, pp. 1273-1280, Dec. 1983.

[24] A. Svensson, "Reduced state sequence detection of partial response continuous phase modulation," *Proc. IEE*, vol. 138, part I, pp. 256-268, Aug. 1991.

[25] P. A. Laurent, "Exact and approximate construction of digital phase modulations by superposition of amplitude modulated pulses," *IEEE Trans. Commun.*, vol. 34, pp. 150-160, Feb. 1986.

[26] J. Huber and W. Liu, "An alternative approach to reduced-complexity CPM-receivers," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 1437-1449, Dec. 1989.

[27] S. J. Simmons, "Simplified coherent detection of CPM," *IEEE Trans. Commun.*, vol. 43, pp. 726-728, Feb./Mar./Apr. 1995.

[28] W. Tang and E. Shwedyk, "A quasi-optimum receiver for continuous phase modulation," *IEEE Trans. Commun.*, vol. 48, pp. 1087-1090, July 2000.

[29] D. E. Gustafson and J. L. Speyer, "Linear minimum variance filters applied to carrier tracking," *IEEE Trans. Auromr. Contr.*, vol. AC-21, pp. 65-73, Feb. 1976.

[30] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Academic Press, 1970.

[31] C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon-limit error-correcting coding and decoding: turbo codes," *Proc. IEEE ICC'93*, pp. 1064-1070, Geneva, May 1993.

[32] B. Benedetto and G. Montorsi, "Iterative decoding of serially concatenated convolutional codes," *Electr. Lett.*, vol. 32, pp. 1186-1188, June 1996.

[33] K. R. Narayanan and G. L. Stüber, "Performance of trellis-coded CPM with iterative demodulation and decoding," *IEEE Trans. Commun.*, vol. 49, pp. 676-687, Apr. 2001.

[34] P. Ho and J. H. Kim, "On pilot symbol assisted detection of CPM schemes operating in fast fading channels," *IEEE Trans. Commun.*, vol. 44, pp. 337-347, March 1996.

[35] R. Balasubramanian, M. P. Fitz, and J. V. Krogmeier, "Pilot symbol aided soft-output demodulation of CPM signal in flat Rayleigh fading channels," in *Proc. Veh. Tech. Conf.*, Pheonix, USA, May 1997, pp. 880-883.

[36] N. Seshadri, "Joint data and channel estimation using blind trellis search techniques," *IEEE Trans. Commun.*, vol. 42, pp. 1000-1011, Feb./March/April 1994.

[37] R. Raheli, A. Polydoros, and C. Tzou, "Per-survivor processing: a general approach to MLSE in unknown environments," *IEEE Trans. Commun.*, vol. 43, pp. 354-364, Feb./March/April 1995.

[38] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284-287, March 1974.

[39] P. Robertson, E. Villebrun and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain," *Proc. IEEE ICC'95*, Seattle, vol. 2, pp. 1009-1013, June 1995.

[40] A. Anastasopoulos and K. M. Chugg, "Adaptive soft-input soft-output algorithms for iterative detection with parametric uncertainty," *IEEE Trans Commun.*, vol. 48, pp. 1638-1649, Oct. 2000.

[41] A. Anastasopoulos and K. M. Chugg, "Adaptive iterative detection for phase tracking in turbo-coded Systems," *IEEE Trans Commun.*, vol. 49, pp. 2135-2144, Dec. 2001.

[42] A. Hansson, K. M. Chugg and T. M. Aulin, "On forward-adaptive versus forward/backward-adaptive SISO algorithms for Rayleigh fading channels," *IEEE Commun. Letters*, vol. 5, pp. 477-479, Dec. 2001.

[43] R. A. Iltis, J. J. Shynk and K. Giridhar, "Bayesian algorithms for blind equalization using parallel adaptive filtering," *IEEE Trans. Commun.*, vol. 42, pp. 10171032, Feb./March/April 1994.

[44] Y. Zhang, M. P. Fitz and S. B. Gelfand, "Soft output demodulation on frequency-selective Rayleigh fading channels using AR channel models," *Proc. IEEE GLOBECOM'97*, Phoenix, pp. 327331, Nov. 1997.

[45] A. Anastasopoulos and A. Polydoros, "Adaptive soft-decision algorithms for mobile fading channels," *Eur. Trans. Commun.*, vol. 9, pp. 183190, March/April 1998.

[46] P. Kovintavewat, J. R. Barry, M. F. Erden and E. M. Kurtas, "Per-survivor iterative timing recovery for coded partial response channels," *Proc. IEEE GLOBECOM'04*, Dallas, vol. 4, pp. 2604-2608, Dec. 2004.

[47] E. L. Kuan and L. Hanzo, "Turbo-coded blind per-survivor processing multiuser detection CDMA," *Proc. IEEE PIMRC'00*, London, vol. 1, pp. 746-750 , Sept. 2000.

[48] J. Heo, K. M. Chugg and A. Anastasopoulos, "A comparison of forward-only and bi-directional fixed-lag adaptive SISOs," in *Proc. IEEE ICC'00*, New Orleans, USA, June, 2000, pp. 1660-1664.

[49] X. Yu and S. Pasupathy, "Innovations-based MLSE for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 43, pp. 1534-1544, Feb./March/April 1995.

[50] B. D. Hart and S. Pasupathy, "Innovations-based MAP detection for time-varying frequency-selective channels," *IEEE Trans. Commun.*, vol. 48, pp. 1507-1519, Sept. 2000.

[51] L.-F. Wei, "Rotationally invariant convolutional channel coding with expanded signal space — Part I: 180°," *IEEE J. Select. Areas Commun.*, vol. 2, pp. 659-671, Sept. 1984.

[52] L.-F. Wei, "Rotationally invariant convolutional channel coding with cxpanded signal space — Part II: Nolinear codes," *IEEE J. Select. Areas Commun.*, vol. 2, pp. 672-685, Sept. 1984.

[53] L.-F. Wei, "Trellis-coded modulation with multidimensional constellations," *IEEE Trans. Inform. Theory*, vol. 33, pp. 483-501, July 1987.

[54] L.-F. Wei, "Rotationally invariant trellis-coded modulations with multidimensional m-PSK," *IEEE J. Select. Areas Commun.*, vol. 7, pp. 1281-1295, Dec. 1989.

[55] S. Benedetto, R. Garello, M. Mondin and G. Montorsi, "Geometrically uniform TCM codes over groups based on L×MPSK constellations," *IEEE Trans. Inform. Theory*, vol. 40, pp. 137-152, Jan. 1994.

[56] M. D. Trott, S. Benedetto, R. Garello and M. Mondin, "Rotational invariance of trellis codes — Part I: encoders and precoders," *IEEE Trans. Inform. Theory*, vol. 42, pp. 751-765, May 1996.

[57] W. Liu and S. G. Wilson, "Rotationally invariant concatenated (turbo) TCM codes," in *Proc. Asilomar Conf. Signals, Systems, Computers*, vol. 1, pp. 32-36, 1999.

[58] R. Nuriyev and A. Anastasopoulos, "Design and robustness analysis of rotationally invariant SCTCM," in *Proc. IEEE ICC'01*, Finland, pp. 2226-2230, June 2001.

[59] R. Nuriyev and A. Anastasopoulos, "Rotationally invariant and iotationally robust codes for the AWGN and the noncoherent channel," *IEEE Trans. Commun.*, vol. 51, pp. 2001-2010, Dec. 2003.

[60] G. Ungerboeck, "Adaptive maximum likelihood receiver for carrier- modulated data-transmission systems," *IEEE Trans. Commun.*, vol. 22, pp. 624636, May 1974.

[61] M. S. Stojanovic, J. G. Proakis and J. A. Catipovic, "Analysis of the impact of channel estimation errors on the performance of a decision-feedback equalizer in fading multipath channels," *IEEE Trans. Commun.*, vol. 43, pp. 877-886, Feb./Mar./Apr. 1995.

[62] X. Tang, M.-S. Alouini and A. J. Goldsmith, "Effect of channel estimation error on M-QAM BER performance in Rayleigh Fading," *IEEE Trans. Commun.*, vol. 47, pp. 1856-1864, Dec. 1999.

[63] L. Cao and N. C. Beaulieu, "Exact error-rate analysis of diversity 16-QAM with channel estimation errors," *IEEE Trans. Commun.*, vol. 52, pp. 1019-1029, June 2004.

[64] Y. Ma, R. Schober and S. Pasupathy, "Effect of channel estimation errors on MRC diversity in Rician fading channels," *IEEE Trans. on Veh. Tech.*, vol. 54, pp. 2137-2142, Nov. 2005.

[65] B. Mielczarek and A. Svensson, "Modeling fading channel-estimation errors in pilot-symbol-assisted systems with application to turbo codes," *IEEE Trans. Commun.*, vol. 53, pp. 1822-1832, Nov. 2005.

[66] M. Dong, L. Tong and B. M. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Trans. Signal Proces.*, vol. 52, pp. 1403-1418, May 2004.

[67] S. Misra, A. Swami and L.Tong, "Optimal training over the Gauss-Markov fading channel: a cutoff rate analysis," *Proc. of IEEE Conf. on Acoust., Speech, and Signal Process.*, May 2004, vol. 3, pp. 809-812.

[68] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, pp. 951-963, April 2003.

[69] H. Vikalo, B. Hassibi, B. M. Hochwald and K. Kailath, "On the capacity of frequncy-selective channels in training-based transmission schemes," *IEEE Trans. Signal Proces.*, vol. 52, pp. 2572-2583, Sept. 2004.

[70] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, pp. 933-946, May 2000.

[71] S. Stein, "Unified analysis of certain coherent and noncoherent binary communications systems," *IEEE Trans. Inform. Theory*, vol. 10, pp. 43-51, Jan. 1964.

[72] W. McGee, "Eye pattern for the binary noncoherent receiver," *IEEE Trans. Commun.*, vol. 19, pp. 634-643, Oct. 1971.

[73] W. Osborne and M. Luntz, "Coherent and Noncoherent Detection CPFSK," *IEEE Trans. Commun.*, vol. 22, pp. 1023-1036, Aug. 1974.

[74] D. Divsalar, M. K. Simon and M. Shahshahani, "The performance of trellis-coded MDPSK with multiple symbol detection," *IEEE Trans. Commun.*, vol. 38, pp. 1391-1403, Sept. 1990.

[75] G. Colavolpe and R. Raheli, "Noncoherent sequence detection," *IEEE Trans. Commun.*, vol. 47, pp. 1376-1385, Sept. 1999.

[76] R.-R. Chen, R. Koetter, U. Madhow and D. Agrawal, "Joint noncoherent demodulation and decoding for the block fading channel: a practical framework for approaching Shannon capacity," *IEEE Trans. Commun.*, vol. 51, pp. 1676-1689, Oct. 2003.

[77] M. C. Cursoy, H. V. Poor and S. Verdu, "The noncoherent rician fading channel-part I: structure of the capacity-achieving input," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2193-2206, Sept. 2005.

[78] N. Jacobsen and U. Madhow, "Code and constellation optimization for efficient noncoherent communication," *Proc. of Asilomar Conf. on Signals, Systems and Computers*, vol. 1, pp. 198-202, Nov. 2004.

[79] D. Raphaeli, G. Sitton and I. Taler, "Geometrically uniform trellis codes for noncoherent detection," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 2182-2188, Nov. 2004.

[80] T. S. Schonhoff, "Symbol error probabilities for M-ary CPFSK: coherent and noncoherent detection," *IEEE Trans. Commun.*, vol. 24, pp. 644-652, Sept. 1976.

[81] M. K. Simon and D. Divsalar, "Maximum-likelihood block detection of noncoherent continuous phase modulation," *IEEE Trans. Commun.*, vol. 41, pp. 90-98, Jan. 1993.

[82] L. Yiin and G. L. Stüber, "Noncoherently detected trellis-coded partial response CPM on mobile radio channels," *IEEE Trans. Commun.*, vol. 44, pp. 967-975, Aug. 1996.

[83] M. K. Simon and M.-S. Alouini, "A unified approach to the probability of error for noncoherent and differentially coherent modulations over generalized fading channels," *IEEE Trans. Commun.*, vol. 46, pp. 1625-1638, Dec. 1998.

[84] G. Colavolpe and R. Raheli, "Noncoherent sequence detection on frequency nonselective slowly fading channels," *Proc. of IEEE Global Telecom. Conf.*, Nov. 1999, pp. 868-872.

[85] G. M. Vitetta, U. Menglai and D. P. Taylor, "An error probability formula for noncoherent orthogonal binary FSK with dual diversity on correlated rician channels," *IEEE Commun. Letter*, vol. 3, pp. 43-45, Feb. 1999.

[86] Q. Zhao and G. L. Stüber, "Joint time and phase recovery for continuous phase modulation," submitted to *IEEE Trans. Commun.* in Feb. 2005 and formally accepted in Jan. 2006.

[87] B. E. Rimoldi, "A decomposition approach to CPM," *IEEE Trans. Inform. Theory*, vol. 34, pp. 260-270, March 1988.

[88] U. Mengali and M. Morelli, "Decomposition of M-ary CPM signals into PAM waveforms," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1265-1275, Sept. 1995.

[89] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. New Jersey: Prentice Hall, 1993.

[90] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. New Jersey: Prentice Hall, 2000.

[91] M. A. Peters and P. A. Iglesias, "A spectral test for observability and detectability of discrete-time linear time-varying systems," *Proc. of the 36th IEEE Conf. on Decision and Control*, San Diego, USA, Dec. 1997, vol. 4, pp. 3920-3925.

[92] B. D. O. Anderson and J. B. Moore, "Detectability and stabilizability of time-varying discete-time linear systems," *SIAM J. Control and Opti.*, vol. 19, pp. 20-32, 1981.

[93] S. Sangsuk-Iam, "Analysis of discrete-time Kalman filtering under incorrect noise covariances," *IEEE Trans. on Auto. Control*, vol. 35, pp. 1304-1309, Dec. 1990.

[94] T. Nishimura, "Modeling errors in Kalman filters," in *Theory and Application of Kalman Filtering*, C. T. Leondes, 1970.

[95] S. S. Saab and G. E. Nasr, "Sensitivity of discrete-time Kalman filter to statistical modeling errors," *Opti. Control Appli. & Methods*, vol. 20, pp. 249-259, Sept./Oct. 1990.

[96] A. N. D'Andrea, U. Mengali and R. Morelli, "The modified Cramer-Rao bound and its application to synchronization problems," *IEEE Trans. Commun.*, vol. 42, pp. 1391-1399, Feb./March/April 1994.

[97] R. Dabora, J. Goldberg and H. Messer, "Inherent limitations in data-aided time synchronization of continuous phase-modulation signals over time-selective fading channels," *IEEE Trans. Signal Process.*, vol. 50, pp. 1470-1482, June 2002.

[98] Q. Zhao, H. Kim and G. L. Stüber, "Innovations-based MAP estimation with application to phase synchronization for serially concatenated CPM," to apprear in *IEEE Trans. Wireless Commun.* (Submitted in Oct. 2003 and formally accepted in March 2005.)

[99] A. Hansson and T. Aulin, "Generalized APP detection for communication over unknown ISI channels," submitted to *IEEE Trans. Commun.*, Dec. 2003.

[100] G. Ferrari, G.Colavolpe and R. Raheli, "On trellis-based truncated-memory detection," *IEEE GLOBECOM'03*, San Francisco, USA, pp. 2218-2222, Dec. 2003.

[101] K. M. Chugg and A. Polydoros, "MLSE for unknown channels–part I: optimality considerations," *IEEE Trans. Commun.*, vol. 44, pp. 836-846, July 1996.

[102] T. A. Summers and S. G. Wilson, "SNR mismatch and online estimation in turbo decoding ," *IEEE Trans. Commun.*, vol. 46, pp. 421-423, April 1998.

[103] A. H. Sayed and T. Kailath, "A state-space approach to adaptive RLS filtering," *IEEE Signal Process. Magazine*, vol. 11, pp. 16-60, July 1994.

[104] Q. Zhao and G. L. Stüber, "Convergence behavior of iterative phase synchronization and detection," *IEEE J. Select. Areas Commun.*, vol. 23, pp. 932-943, May 2005. (Submitted in March 2004 and formally accepted in Nov. 2004.)

[105] P. Hoeher and J. Lodge, "Turbo DPSK: Iterative differential PSK demodulation and channel decodings," *IEEE Trans Commun.*, vol. 47, pp. 837-843, June 1999.

[106] D. Divsalar, S. Dolinar and F. Pollara, "Iterative turbo decoder analysis based on density envolution," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 891-907, May 2001.

[107] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, pp. 1727-1737, Oct. 2001.

[108] M. M. Meerschaert and H. Scheffler, *Limit Distributions for Sums of Independent Random Vectors: Heavy Tails in theory and practice*. New York: Wiley, 2001.

[109] G. Ferrari, A. Anastasopoulos, G.Colavolpe and R. Raheli, "Adaptive iterative detection for the phase uncertain channel: limited-tree-serach versus truncated-memory detection," *IEEE Trans. Veh. Tech.*, vol. 53, pp. 433-442, Mar. 2004.

[110] C. He, A. Banerjee, D. J. Costello and P. C. Massey, "On the performance of low density multiple turbo codes," in *Proc. 40th Annual Allerton Conf. on Commun., Control, and Comput.*, Monticello, IL, Oct. 2002.

[111] M. Moeneclaey, S. Starzak and H. Meyr, "Cycle slips in synchronizers subjected to smooth narrow-band loop noise ," *IEEE Trans. Commun.*, vol. 37, pp. 867-874, June 1988.

[112] G. De Jonghe and M. Moeneclaey, "Asymptotic cycle slip probability expression for the NDA feedforward carrier synchronizer for M-PSK," *IEEE ICC'93*, Geneva, pp. 502-506, May, 1993.

[113] Q. Zhao and G. L. Stüber, "Turbo synchronization for serially concatenated CPM," to appear in *IEEE ICC'2006*.

[114] Y. Sung and L. Tong, "Blind channel tracking for long-code WCDMA with linear interpolation model," in *The 36-th Asilomar Conf. on Signals, Systems and Computers*, vol. 1, pp. 694-698, Nov. 2002.

[115] F. M. Gardner, "Interpolation in digital modems— Part I: Fundamentals," *IEEE Trans. Commun.*, vol. 41, pp. 501-507, March 1993.

[116] L. Erup, F. M. Gardner and R. A. Harris, "Interpolation in digital modems—Part II. Implementation and performance," *IEEE Trans. Commun.*, vol. 41, pp. 998-1008, June 1993.

[117] J. G. Proakis, *Digital Communications*. McGraw-Hill, 4ed, 2000.

[118] Q. Zhao and G. L. Stüber, "Noncoherent and coherent sequence detection on Rayleigh fading channels," submitted to *IEEE Trans. Commun.* in Feb. 2006.

[119] G. L. Stüber, *Principles of Mobile Communication*. Boston: Kluwer Academic Publishers, 2ed, 2001.

[120] M. Schwartz, W. R. Bennett and S. Stein, *Principles of Mobile Communication*. New York: McGraw-Hill, 2ed, 1996.

[121] G. L. Turin, "The characteristic function of hermitian quadratic forms in complex normal variables," *Biometrika*, vol. 47, pp. 199-201, June 1960

[122] G. L. Turin, "On optimal diversity reception," *IRE Trans. Inform. Theory*, vol. IT-7, pp. 154-166, July 1961.

[123] J. N. Pierce and S. Stein, "Multiple diversity with nonindependent fading," *Proc. of IRE*, vol. 48, pp. 89-104, Jan. 1960.

[124] J. E. Smee and N. C. Beaulieu, "Error-rate evaluation of linear equalization and decision feedback equalization with error propagation," *IEEE Trans. Commun.*, vol. 46, pp. 656-665, May 1998.

[125] V. Y. Krachkovsky and Y. X. Lee, "Error propagation evaluation for MDFE detection," *IEEE Trans. Magnetics*, vol. 33, pp. 2770-2772, Sept. 1997.

[126] Q. Zhao and G. L. Stüber, "Joint time and phase recovery for CPM and its asymptotic behavior," *IEEE GlobeCom'2005*, St'Louis, U.S.A, pp. 2302-2306, Dec. 2005.

[127] D. Makrakis and K. Feher, "Multiple differential detection of continuous phase modulation signals," *IEEE Trans. Veh. Tech.*, vol. 42, pp. 186-196, May 1993.

[128] G. Colavolpe and R. Raheli, "Noncoherent sequence detection of continuous phase modulations," *IEEE Trans. Commun.*, vol. 47, pp. 1303-1307, Sept. 1999.

[129] J. Wu, C. Xiao and J. C. Oliver, "Time-varying and frequency-selective channel estimation with unequally spaced pilot symbols," in *Proc. IEEE ICASSP'03*, Hong Kong, vol. 4, pp. 620-623, April 2003.

[130] J. K. Cavers, "An Analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Tech.*, vol. 40, pp. 686-693, Nov. 1991.

[131] P. Hoeher and F. Tufvesson, "Channel estimation with superimposed pilot sequence," *Proc. IEEE GLOBECOM'99*, Rio de Janeiro, vol. 4, pp. 2162-2166, Dec. 1999.

[132] S.-C. Hong, J.-S. Joo and Y. H. Lee, "Per-survivor processing sequence detection for DS/CDMA systems with pilot and traffic channels," *IEEE Commun. Letters*, vol. 5, pp. 346-348, Aug. 2001.

[133] R. J. Piechocki, C. Andrieu, A. R. Nix and J. P. McGeehan, "Joint blind and semi-blind detection and channel estimation for space-time trellis coded systems," *Proc. IEEE ICASSP'03*, Hong Kong, vol. 6, pp. 349-352, April 2003.

[134] E. Baccarelli and R. Cusani, "Combined channel estimation and data detection using soft statistics for frequency-selective fast-fading digital links," *IEEE Trans. Commun.*, vol. 46, pp. 424427, April 1998.

[135] M. K. Simon and C. C. Wang, "Differential detection of gaussian MSK in a mobile radio environment," *IEEE Trans. Veh. Tech.*, vol. 33, pp. 307-320, Nov. 1984.

VITA

Qing Zhao received his B.S.E.E and M.S.E.E degrees both in electrical engineering from Tsinghua University, Beijing, China, and New Mexico State University, Las cruces, USA, respectively. Since the summer of 2002 he started to work toward the Ph.D degree in the School of Electrical and Computer Engineering at Georgia Institute of Technology, Atlanta, USA. During the summers of 2004 and 2005, he worked in the Sector of Mobile Device at Motorola Inc. on developing and analyzing link simulators for Beyond 3G cellular communication systems. Now he is working in the DSP group at Marvell Semiconductor Inc. as a senior wireless system design engineer. His research interests include the areas of communication and signal processing, specified in the algorithm design and analysis for iterative demodulation/decoding with parametric uncertainty, and its application to synchronization, channel estimation, interference suppression and multiuser detection.